

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:27:40 ; Search time 17 Seconds

(without alignments)  
1914.344 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 3313  
Sequence: 1 MSSRAGPMGKKNVGGYRP.....EPDEALQMLYRKHLNEDD 625

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	196	5.9	2130 1	BA2B_CHICK
2	190	5.7	1360 1	CING_XENLA
3	183	5.5	1191 1	CING_MOUSE
4	180	5.4	1972 1	BA2B_HUMAN
5	178	5.4	1197 1	CING_HUMAN
6	174	5.3	1937 1	MYH8_HUMAN
7	173.5	5.2	978 1	RAS0_AQUAE
8	173	5.2	1084 1	MYSS_RABIT
9	172.5	5.2	593 1	TARA_HUMAN
10	172.5	5.2	1407 1	TRHY_RABIT
11	171.5	5.2	2116 1	MYSS_DICDI
12	169	5.1	1938 1	MYH4_RABIT
13	167.5	5.1	1898 1	TRHY_HUMAN
14	166.5	5.0	1938 1	MYHD_HUMAN
15	165	5.0	724 1	HMHR_HUMAN
16	165	5.0	919 1	INCE_HUMAN
17	165	5.0	2663 1	CENE_HUMAN
18	163.5	4.9	1939 1	MYH1_HUMAN
19	163	4.9	1756 1	PEPL_HUMAN
20	163	4.9	1790 1	USO1_YEAST
21	162	4.9	9660 1	STKA_MOUSE
22	162	4.9	1969 1	MYSA_CAEEL
23	161.5	4.9	1875 1	MLP1_YEAST
24	161.5	4.9	3911 1	AKA9_HUMAN
25	161	4.9	1102 1	MYSC_CHICK
26	160	4.8	1972 1	MYHB_HUMAN
27	159.5	4.8	1549 1	TRHY_SHEEP
28	159.5	4.8	1556 1	PROS_DROVI
29	159.5	4.8	1938 1	MYH6_RAT
30	159.5	4.8	2230 1	GOM4_HUMAN
31	158.5	4.8	1979 1	TRIA_HUMAN
32	158	4.8	1978 1	MYHB_CHICK
33	157.5	4.8	1941 1	MYH2_HUMAN

34	157.5	4.8	1972 1	MYHB_MOUSE	008638 mus musculus
35	156.5	4.7	976 1	SCPI_HUMAN	015431 homo sapien
36	156.5	4.7	1755 1	PEPL_HUMAN	09269 mus musculus
37	156	4.7	1169 1	SMC_MERJA	059037 methanococc
38	155.5	4.7	1427 1	REST_HUMAN	030622 homo sapien
39	155	4.7	1020 1	CF60_HUMAN	08nd25 homo sapien
40	155	4.7	1085 1	YAF4_SCHPO	009863 schizosacch
41	154.5	4.7	1972 1	MYHB_RABIT	035748 oryctolagus
42	154.5	4.7	2363 1	SPCO_MOUSE	062261 mus musculus
43	154	4.6	852 1	RAS0_THEMA	091x11 thermotoga
44	154	4.6	1939 1	MYH6_HUMAN	013533 homo sapien
45	154	4.6	1940 1	MYH3_CHICK	002565 gallus gall

## ALIGNMENTS

RESULT 1  
ID BA2B\_CHICK STANDARD; PRT; 2130 AA.  
AC Q9DEI3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Bromodomain adjacent to zinc finger domain 2B (Extracellular matrix protein F22).  
GN BA2B.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yoon H., Philip N.D.;  
RT "Cloning of a new extracellular matrix protein expressed in retina."  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBS databases.  
CC - FUNCTION: May play a role in transcriptional regulation.  
CC - SIMILARITY: Belongs to the MAL family.  
CC - SIMILARITY: Contains 1 bromodomain.  
CC - SIMILARITY: Contains 1 DDT domain.  
CC - SIMILARITY: Contains 1 methyl-binding (MBD) domain.  
CC - SIMILARITY: Contains 1 PHD-type zinc finger.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF224275; AAC36791.1; -.  
DR HSSP; Q92831; 1B91.  
DR InterPro; IPR001487; Bromodomain.  
DR InterPro; IPR004022; DDT domain.  
DR InterPro; IPR001739; Methyl-CpG\_bind.  
DR InterPro; IPR001965; ZnF\_PHD.  
DR Pfam; PF00439; bromodomain; 1.  
DR Pfam; PF02791; DDT; 1.  
DR Pfam; PF01429; MBD; 1.  
DR Pfam; PF00628; PHD; 1.  
DR PRINTS; PR00503; BROMODOMAIN.  
DR SMART; SM00297; BROMO; 1.  
DR SMART; SM00571; DDT; 1.  
DR SMART; SM00249; PHD; 1.  
DR PROSITE; PS00633; BROMODOMAIN\_1; FALSE\_NEG.  
DR PROSITE; PS50016; ZF\_PHD\_2; 1.  
DR PROSITE; PS01359; ZF\_PHD\_1; FALSE\_NEG.  
DR PROSITE; PS50016; ZF\_PHD\_2; 1.  
DR Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;  
KW Nuclear protein; DNA-binding.

```

FT DOMAIN 693 743 MBD.
FT DOMAIN 1010 1075 DDT.
FT ZN PING 1895 1945 PHD-TYPE.
FT DOMAIN 2039 2109 BROMODOMAIN.
FT DOMAIN 9 46 SER-RICH.
FT DOMAIN 184 208 SER-RICH.
FT DOMAIN 212 219 POLY-GLU.
FT DOMAIN 544 610 ASP-RICH.
FT DOMAIN 797 984 COILED COIL (POTENTIAL).
FT DOMAIN 882 886 POLY-LYS.
FT DOMAIN 1219 1262 ASP-RICH.
FT DOMAIN 1909 1912 POLY-LEU.
SQ SEQUENCE 2130 AA; 236140 MW; 208C48FB0BA68F70 CRC64;

Query Match 5.9%; Score 196; DB 1; Length 2130;
Best Local Similarity 20.9%; Pred. No. 0.01;
Matches 144; Conservative 96; Mismatches 262; Indels 186; Gaps 28;

QY 19 RPEVEQLVGLAGRLASSQDDGEMEVISKKNKPKGNTSG-----KTVWSONSNP--P 71
DB 374 RSKREQYKQTPPAQLKKQESSKNLKVIVSLSSKPTSCSPAHQKLTSLNNHSPFLT 433
QY 72 RAMGGQOQGRG-----SNVSGRGNNVSGHNGNGKGIQANISGRRLS 115
DB 434 NALLGNHPNGVIGSVIGVPLALTTKQKQTKINESVAIASSTPESLPVNLIS---ACG 489
QY 116 RKYDNNFVAPPVSRPPELEGW-----NMQARGSAQH-----TAVCEFPVY----- 157
DB 490 KKTGNRTLVVPSYSPVLPFGSGKDKPVSNNAVNAVKTOHCLPSAKLVVEGFRGVSDAPG 549
QY 158 -----EDVDNASEENDSDALDSDDDLA-----SDDYSDVS 191
DB 550 SKESDSDSDDDDDDDDDDDDDDD--DSDSDQSESDSDSDSDSDSDSDSDSDSDSDSDS 607
QY 192 QKSHGRKQKWKFKFPGSLDLSLEQI--NEPQKQMTCPACQNGPGLDWNHPLPL-- 247
DB 608 DTDTEGKPLTKTKTGSSIKSSSGPVASHPTPLNQVAKTPSSAPSL--CEPTQPAVFL 666
QY 248 -----AHARTGARVRLHRELAEVLEKD-----LQMGASVIRPG 284
DB 667 GTTPSTLTSSHSGISKRKRVYDERELRYPLEYQWQRETRINFGRLQGEVAVFAPGCK 726
QY 285 -----IYG-QWKLGEDEKDYELVWPMPVITIMTRLDKDDNKKMLGMNGQLEFYD 335
DB 727 KLRQYEVVGVQWCLL-----KEEEVY--FCIRAMEGRGRPPNP-----DQGH 769
QY 336 KYEALPARHSYQPGHGRGMSVLMFESSA--TGYLEARLHRELAEMGLDRIAMGQKSMF 393
DB 770 SREESRRRRKGRPPNVGSTEPLDSDALKRLQLQOETAFOAQIKLLRKLQKOEQARA 829
QY 394 SGGVROLVGFPLATKODLIDFNHOSQKTRIKFELK--SYDEMVKELROISEDN-----OOL 448
DB 830 AKEKKQAQALMAAE-----KROKEQIKIMKQOEKIKRLOQIRMEKEILRAQOI 878
QY 449 NYFNKLISKQNAKAVIESELEIMSEKLR-----TADNRIVYQRTQWQHEQNRREMDAH 504
DB 879 LEAKKKKKKEAANKLLEAKRKIKEMKROQAVLKHQOLEHRRIDMERRRQUMML- 937
QY 505 DRFMDISIKOIHERRDAKENFEMLOOQERAKVVGQOONINPSSNDCKRAAEVSSFI 564
DB 938 -----MKAMEAKKAEK--EHLKQEKR-----DEKRLN----- 964
QY 565 EFQEKEMEEVEEREMLIKQDEKQMEDM 592
DB 965 --KERKLEFORLELEM-AKELKKQMEDM 989

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cingulin.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Oocyte;
RX MEDLINE=20082893; PubMed=10613913;
RA Cordenonci M., D'Attri F., Hammar E., Parry D.A.D., Kenrick-Jones J.,
RT "Cingulin contains globular and coiled-coil domains and interacts with
RT ZO-1, ZO-2, ZO-3 and myosin."
RL J. Cell Biol. 147:1569-1581(1999).
RN [2]
RP FUNCTION IN TIGHT JUNCTION BIOGENESIS.
RX MEDLINE=97029068; PubMed=8875080;
RA Cardellini P., Davanzo G., Citi S.;
RT "Tight junctions in early amphibian development: detection of
RT functional cingulin from the 2-cell stage and its localization at the
RT boundary of distinct membrane domains in dividing blastomeres in low
RT calcium."
RL Dev. Dyn. 207:104-113(1996).
RN [3]
RP FUNCTION IN TIGHT JUNCTION BIOGENESIS.
RX MEDLINE=20400099; PubMed=10940624;
RA Fesenko I., Kurth T., Sheth B., Fleming T.P., Citi S., Hausen P.;
RT "Tight junction biogenesis in the early Xenopus embryo."
RL Mech. Dev. 96:51-65(2000).
RN [4]
RP INTERACTION WITH OCCCLUDIN.
RX MEDLINE=99421641; PubMed=10491082;
RA Cordononci M., Turco F., D'Attri F., Hammar E., Martinucci G.,
RA Meggio F., Citi S.;
RT "Xenopus laevis occludin. Identification of in vitro phosphorylation
RT sites by protein kinase CK2 and association with cingulin."
RL Eur. J. Biochem. 264:374-384(1999).
RN [5]
RP INTERACTION WITH F-ACTIN.
RX MEDLINE=21538627; PubMed=11682052;
RA D'Attri F., Citi S.;
RT "Cingulin interacts with F-actin in vitro."
RL FEBS Lett. 507:21-24(2001).
RN [6]
RP INTERACTION WITH ZO-1.
RX MEDLINE=22140336; PubMed=12023291;
RA D'Attri F., Nadalutti F., Citi S.;
RT "Evidence for a functional interaction between cingulin and ZO-1 in
RT cultured cells."
RL J. Biol. Chem. 277:27757-27764(2002).
RN [7]
RP FUNCTION: Probably plays a role in the formation and regulation of
RN the tight junction (TJ) paracellular permeability barrier,
RN possibly by linking ZO proteins to the actomyosin cytoskeleton.
RN -1- FUNCTION: Parallel homodimer (by similarity). Binds ZO-1 and ZO-2
RN in vivo, and ZO-3, myosin and occludin in vitro, possibly
RN directly. Acts as an F-actin bundling protein in vitro.
RN -1- TISSUE SPECIFICITY: Localized on the cytoplasmic face of tight
RN junctions of polarized epithelia and some endothelia.
RN -1- DEVELOPMENTAL STAGE: A maternally synthesized protein. Found in
RN the apical cortex in the fertilized egg, where it is associated
RN with cytoskeleton filaments, it is recruited to tight junctions
RN before ZO-1 and occludin. Nascent tight junctions are in place
RN by the two-cell stage.
RN -1- DOMAIN: Deletion of the ZO-1 interaction motif (ZIM) decreases but
RN does not abolish colocalization with ZO-1.
RN
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL Outstation
RN at the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF207901; AAF20208.1; ALT\_INIT.  
 DR HSSP; P03034; ILMB.  
 DR GO; GO:0005923; C:light junction; NAS.  
 DR GO; GO:0003779; F:actin binding; IDA.  
 KW Tight junction; Coiled coil.  
 FT DOMAIN 1 427 HEAD  
 FT DOMAIN 428 1322 COILED COIL (POTENTIAL).  
 FT DOMAIN 1323 1360 TAIL.  
 FT DOMAIN 88 261 SER-RICH.  
 FT SITE 1 370 INTERACTS WITH ZO-3 AND MYOSIN.  
 FT SITE 33 47 ZIM.  
 FT SITE 93 266 INTERACTS WITH F-ACTIN.  
 FT SITE 142 287 INTERACTS WITH ZO-2.  
 FT SITE 369 1360 INTERACTS WITH MYOSIN.  
 SQ SEQUENCE 1360 AA; 158465 MW; 9066CA572401B7B0 CRC64;

Query Match 5.7%; Score 190; DB 1; Length 1360;  
 Best Local Similarity 20.8%; Pred. No. 0.013;  
 Matches 152; Conservative 110; Mismatches 242; Indels 226; Gaps 32;

QY 2 SSR--AGPMSKKNVGGYRPEVEQLVQGLAGTRLASQDDGGEMWISKKNNKNGNTS 59  
 DB 194 SSMDSMWSSSQKKNQDPSL-----GYSATSSQSTST--VSNKTKKNGLSTS 241  
 QY 60 GKTWVSQNSNPPRANG-----GQOGRGSNVN-----GRGNVNSGRGN--- 97  
 DB 242 SPSNQNMEDIDTKPLSSVSLINKPKIGQVGRTRARRSQUALDKERRKSSQLGRKNYHD 301  
 QY 98 -GNGRIGQANISGRK-----ALSRYDNMFVAPPVSRPPLLEGMMQWARG-----S 145  
 DB 302 TADSRRIIVEKQNEVQTMREPVNASNRSPRQTLERGDISTKTLTKEMLDQDEEPIVILK 361  
 QY 146 ACHTAQVEFPDVEDVNASEENDSDALDDSDDDASDDYDVSQKSHGRKNKWFK 205  
 DB 362 QGRTVQSEF-----QKSTPDLARDQPD-----GSDPTR-----E 392  
 QY 206 KPPGSLDSISIEQINEPORQ-----HCPACQNGPQALDWNVNLHPLLAHRTYGARRVK 260  
 DB 393 MTFGILREGSLSESENTLRKKTSTLEKLPSTLVQVQPE-DTISL-----GSOKEKEL 441  
 QY 261 HRELAIV-----LEKDLQMRASVTPGCEIYGCKKGLG 293  
 DB 442 ERKVALQRLQDDEMKQKMLKTSQGRPYAKQWRLTELESEKESRLEKEL----- 494  
 QY 294 EDEKDEIWPMPVVIIMNTRLDKDDNDKMLGMQNOELLEFDKYKALRAHSHYQPOGHRG 353  
 DB 495 --EKKKNELSSAMQELMEVRMGKEQVETKRTMEDLMSKELSHLRKAGTSPD---K 549  
 QY 354 MSVLMESSATGYLEARLHRELAENGDLRIANGQKRSMSFGVROLYGFPLATKQDLDIF 413  
 DB 550 LALLK-----ELEEVQDELDEV-----LQIRQ-----KOE-ELL 577  
 QY 414 NQHSQKTKLKFELKSYQEMVVEVLEQISB-----DNQOL-----NYFKNLSKONHAKY 464  
 DB 578 RQKDRRLTALKGALKQEVANHDLDVRQYQNDVQOLRKNNDNVSQDLSLESSERQK 637  
 QY 465 -----LESLSEIMS-----EKLRTAEDNRIVRQ-----RTMQHSEONNEEM-DADHDF 507  
 DB 638 NQVVRRLQRELESSSPDISQWKMFPQKNEBELRSTQELLQMLLEKESBDELKTRDRF 697  
 QY 508 FM-----DSIKQIHERRDAKEENFEMLOOQER-AK 536  
 DB 698 SLQSELAQVKGSVDPGEVASVRKELQRYODQLKLSVDKQKVEEN---LQOREREMSA 754  
 QY 537 VVQGOQONINPSSNDCKRAEVSSEFIEQEKEMEFVBERMLKQOKKMKEDKKRH 596  
 DB 755 LKQTLKEEVSQGRDETIVRLR-EQLQSEVVMVKKENGLAKESRR-IQDQKQVLLERKORH 812  
 QY 597 HREIFDLERE 606

DB 813 EETVHQERE 822

RESULT 3  
 CING MOUSE STANDARD; PRT; 1191 AA.  
 ID CING MOUSE  
 AC P59242;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Cingulin.  
 GN CGN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mallory S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalski U., Smalins D.E.,  
 RA Schenker A., Schein J.E., Jones S.J.M., Maize M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: Probably plays a role in the formation and regulation of  
 the tight junction (TJ) paracellular permeability barrier (By  
 similarity).  
 CC -1- SUBUNIT: Homodimer. Interacts with TJP1/ZO-1 (By similarity).  
 CC -1- DOMAIN: Deletion of the ZO-1 interaction motif (ZIM) decreases but  
 does not abolish colocalization with ZO-1 (By similarity).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 5.5%; Score 183; DB 1; Length 1191;  
 Best Local Similarity 18.9%; Pred. No. 0.025;  
 Matches 171; Conservative 126; Mismatches 281; Indels 326; Gaps 38;

15 OCGYRPE-----VEOLVOGLAGTR--LASQDDGEMEWISKONKPKGTSCKTW 63  
 31 RSGRRPARDARASTYGVAVRVLGIGAGFPVVLNISEKDTDFGVQIKGNNRSGPGA--- 87  
 64 VQNSNPBPAMGQOQGGNSVSGRNNVSGRNGNGI-----QANISG-----RG 111  
 88 LSSDSLSPENPYQVGRPATRSSTSDPEKDLNGLKINSQSQASLTGLAFMSPNRS 147  
 112 RALSRTYDNNFVAPRVS-----RPPV-----EGGNNWQARGSGAHTAVQ 152  
 148 TSLLE-----LAPKPTSSINTIDTAPLSVDSLKPKDSQGG---QVRGRTGRT--R 196  
 153 EPPVEDVDVNAEENSDALDDSDDLAS-----DDYDSVQSCKSGSRKONWPFKF 207  
 197 TLPHQCKRQSQSLDRPLRPTREERHQSANWTRGTKYDNHVSCKNSQSQSP-FSSF 255  
 208 FGSIDS-----LSIEQINER-----ORWHCPACONGPAGADWY----- 241  
 256 SRSRQTDWVLQSFETRDPAWQFSTPDLADORE-----TAPPGADHVKATYGI 309  
 242 -----NHPV-----LAHARTKG-ARRVYKJHRLAV----- 267  
 310 LREBSSEASVRRKVSIVLEQWPLGAVSPASTKALAGQALTRKMBELQKKLDEYVK 369  
 268 -----LEKLOMRGASVIRPCGRIYQWKG-----LGEDE----- 296  
 370 RQKLEPSRVGLEROLEEKAEECHRLQBELERRKGEVQSSKEIQNMKLLGQEGRLHGL 429  
 297 -----KDY-----EIVWPWYIIMNTFLDDDDDKML 323  
 430 ENQVKELOLKLKHSQPSGKESILLKDLDTRELLBELLEKORVSEQLREHETALK 489  
 324 GNGNOELLEFPKYALAAHSYG-POGRHGSVLMFESSATGYLEAER-----LH 373  
 490 GALKKEEVASHDOVEHVRLOQYQRTQEURRSMQDATTQDHA--LEAERQMSLVRELQ 546  
 374 RELAEWGLDRIMGQRKSMFSGVRLYQVFLATKQDDIFMHSQCKTRLKFKLSYQEM 433  
 547 RELIEFSEETGHW--QSWFQKKEEL--RATKQEL--LOLRMEKEMEELGEGKMEV 597  
 434 VKELEROI--SEDOQOLNFPKKL-----SKONKHA-----KYLESELEIM 472  
 598 LQDLDLQARASTDTQVVELEKKEKRLRTQGELEKELQAEQONQVGRHNRVLEKQALAL 657  
 473 SEKL--RRTAENRIRYORTKMOHEQN-----REEM 501  
 658 REEDRGRLEQONLOLQTLQRLQDCEASAKAVASETEAMMLGORATVETTLRTQ 717  
 502 DAHDF-----FMDSIKOIHBRDAKEENFEMLOOE- 533  
 718 EENDEFRRILGLEQLKEARGLAEGEAEVARELARKVHLEVEKQQLFEALMAAQEBG 777  
 534 -----RAKVV-----GOOQONIPSSNDCKRAEVSFTIEOE----- 568  
 778 NLAALAKLLEVRLEDAORGLARLQGEQOLNRLALEEGKQREMLRSKLEBOKRLNLR 837  
 569 -----KEMEFVEEREMLIKQCKKMDKKHHEIFLEKEF--DEALEQUMYKGLH 621  
 838 TVDRLEKLEQIGDSSKLALQOLQOQWEDYKARKEVADARQADKMSAEKNSGGLS 897  
 622 NEDD 625  
 898 RLQD 901

GN BAZ2B OR KIA1476.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxId=9606;  
 RN [1] SEQUENCE FROM N.A. (ISOFORM 1).  
 RP TISSUE=Testis;  
 RC MEDLINE=20130112; PubMed=1062543;  
 RX Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.,  
 RA Jones M.H., Hamana N., Nezu J., Shimane M.,  
 RT "A novel family of bromodomain genes.";  
 RL Genomics 63:40-45(2000).  
 RN [2] SEQUENCE FROM N.A. (ISOFORM 3).  
 RP Ohara O., Nagase T., Kikuno R.,  
 RA Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 RN [3] SEQUENCE OF 753-1972 FROM N.A.  
 RP MEDLINE=20277482; PubMed=10819331;  
 RX Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.,  
 RT "Prediction of the coding sequences of unidentified human genes. XVII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 7:143-150(2000).  
 RN [4] SEQUENCE OF 1-776 FROM N.A. (ISOFORM 1).  
 RP TISSUE=Melanoma;  
 RC Anorgue W., Winkner U., Mewes H.-W., Weil B., Wiemann S.,  
 RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [5] SEQUENCE OF 1-799 FROM N.A. (ISOFORM 2).  
 RP TISSUE=Skeletal muscle;  
 RC MEDLINE=22388257; PubMed=12477932;  
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Bhat N.K.,  
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Houghton R., Marutani M., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Diatchenko L., Soares M.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Staphenko M., Ustin T.B., Toshiki S., Adami R.D., Mulhally S.J.,  
 RA Brownstein M.J., Loggiano N.A., Peters G.J., Adams R.D., Mulhally S.J.,  
 RA Raha S.S., McManus P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Boshak S.A., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliyil S.W.,  
 RA Richards S., Murray D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Vallian D.K., Muzny D.M., Ketterman M., Madan A., Rodriguez A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez A.,  
 RA Whiting M., Madan A., Young A.C., Green E.D., Dickson M.C.,  
 RA Blakesley R.W., Touchman J.W., Schmutz J., Myers R.M.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Small D.E.,  
 RA Schmechel A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [6] SEQUENCE OF 427-1215 FROM N.A. (ISOFORM 3).  
 RP Itogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,  
 RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Ishii S., Kawai Y.,  
 RA Saito K., Yamamoto U., Wakamatsu A., Nakamura Y., Nagahara K.,  
 RA Masuko Y., Kanehori K.,  
 RT "NEBO human cDNA sequencing project.";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 RN [7] SEQUENCE OF 1524-1972 FROM N.A.  
 RP TISSUE=Testis;  
 RC Pouetka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.,  
 RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: May play a role in transcriptional regulation  
 CC interacting with ISM1.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=3;  
 CC Comment-Experimental confirmation may be lacking for some  
 CC isoforms;



CC Name=1;  
CC IsoId=Q9UIF8-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9UIF8-2; Sequence=VSP\_000553;  
CC Name=Inferred from Ref.4;  
CC Name=3;  
CC IsoId=Q9UIF8-3; Sequence=VSP\_000554;  
CC Note=Inferred from Ref.5;  
CC TISSUE SPECIFICITY: Expressed at varying levels in several  
CC tissues, whereas a smaller transcript was expressed specifically  
CC in testis.  
CC -1- SIMILARITY: Belongs to the WAL family.  
CC -1- SIMILARITY: Contains 1 bromodomain.  
CC -1- SIMILARITY: Contains 1 DDT domain.  
CC -1- SIMILARITY: Contains 1 methyl-binding (MBD) domain.  
CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.  
CC -1- CAUTION: Ref.5 sequence differs from that shown due to a  
CC frameshift in position 731.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
CC -----  
DR EMBL; AB032355; BA869212.1; -  
DR EMBL; AB040909; BA96000.2; ALT\_INIT.  
DR EMBL; AL834381; CAD39044.1; ALT\_INIT.  
DR EMBL; BC012576; AA112576.1; ALT\_FRAME.  
DR EMBL; AK027612; BAB5231.1; ALT\_INIT.  
DR EMBL; AL080173; CAB45759.1; -  
DR PIR; T12495; T12495.  
DR HSSP; Q92831; 1B91.  
DR GeneW; HGNC:963; BAZ2B.  
DR MIM; 605683; -  
DR InterPro; IPR001487; Bromodomain.  
DR InterPro; IPR004023; DDT dom.  
DR InterPro; IPR001739; Methyl\_CpG\_bind.  
DR InterPro; IPR001965; ZnF\_PHD.  
DR Pfam; PF00439; bromodomain; 1.  
DR Pfam; PF02791; DDT; 1.  
DR Pfam; PF01429; MBD; 1.  
DR Pfam; PF00628; PHD; 1.  
DR PRINTS; PRO0503; BROMODOMAIN.  
DR SMART; SM00297; BROMO; 1.  
DR SMART; SM00391; MBD; 1.  
DR SMART; SM00249; PHD; 1.  
DR PROSITE; PS00633; BROMODOMAIN\_1; FALSE\_NEG.  
DR PROSITE; PS50014; BROMODOMAIN\_2; 1.  
DR PROSITE; PS50827; DDT; 1.  
DR PROSITE; PS01359; ZF\_PHD\_1; FALSE\_NEG.  
DR PROSITE; PS50016; ZF\_PHD\_2; 1.  
KW Transcription regulation; Bromodomain; Zinc-finger; Coiled coil;  
KW Nuclear protein; DNA-binding; Alternative splicing.  
FT DOMAIN 549 601  
FT DOMAIN 891 956  
FT ZN\_FING 1735 1785  
FT DOMAIN 1881 1951  
FT DOMAIN 45 69  
FT DOMAIN 73 79  
FT DOMAIN 399 470  
FT DOMAIN 641 672  
FT DOMAIN 687 685  
FT DOMAIN 706 875  
FT DOMAIN 1100 1143  
FT DOMAIN 1138 1179  
FT VARSPPLIC 437 534  
FT VARSPPLIC 593 626

FT CONFLICT 137 137 S -> F (IN REF. 1).  
FT CONFLICT 226 226 S -> L (IN REF. 1).  
FT CONFLICT 627 627 G -> E (IN REF. 1).  
FT CONFLICT 722 722 E -> K (IN REF. 5).  
FT CONFLICT 790 799 RIKEKEMRO -> QKKKKKKKK (IN REF. 5).  
FT CONFLICT 1128 1128 K -> Q (IN REF. 1).  
FT CONFLICT 1183 1183 Q -> P (IN REF. 1).  
FT CONFLICT 1195 1195 Q -> R (IN REF. 1).  
FT CONFLICT 1453 1453 L -> S (IN REF. 3).  
FT CONFLICT 1838 1838 K -> Q (IN REF. 1).  
SQ SEQUENCE 1972 AA; 220709 MW; DAF105C927B91569 CRC64;  
  
Query Match 5.4%; Score 180; DB 1; Length 1972;  
Best Local Similarity 21.0%; Pred. No. 0.063;  
Matches 114; Conservative 83; Mismatches 189; Indels 158; Gaps 21;  
  
QY 140 QARQSAQHTAVQEPFVDDVDVNAEEENDSDALDDSDDLAS----- 183  
DB 394 QFRGTDSDIPSSKDSNSDEDEDEDEDEDEDEDEDEDESDQSSEDSNSSESDTEGSEER 453  
QY 184 --DDVSDVSQKSHGRKONKMKFFGSL--DSLSEIQINERQRMHCPACONGGAI 239  
DB 454 DDDDKQDDSDSDTESEKTSMKLNTTSVKSFSMSLTGHSFRLHIAKAPGAPALC 513  
QY 240 WYMLHPL-----AAATKGAARVKLHRELAFLVKD-----LQMR 275  
DB 514 SESQSAFLGTSSTLTSSPHSSTSKRRVTDRELRIPLEYQMGRETRIRNFCGRLOGE 573  
QY 276 GASVIFCGEITYGQ-----KGLGEDEKD-----YEW-----IYW--- 303  
DB 574 VAVYAPCGKRLRQYPPVVIKYLNRNGIMDISRDNFSAKIRVGDPEARDGPGOMQWCL 633  
QY 304 -----PPWVIMNTRLDKQNDNDLWGMQGELEFYEDK---BALARHSYSGQGRGMS 355  
DB 634 KEEDVLPRIAMEGRGRPPNP-----DRQAREESMRRKRGPPVNGNA 679  
QY 356 VLMFESSA--TGYLEARLHRELAEMGLDRIAMQGRSMFGGVRQLYGLATKODLDF 413  
DB 680 EPLDNDADKLRLQAOELARQAOIKLRKQKQXQAVAKAKKQAOIAMEE----- 734  
QY 414 NQHSQKTRIKPELK-SYQEMVVKELRQISEDN---QQLNYPKNLSKQNKRIAYLEES 468  
DB 735 -----KRXKEQIKIMKQOEKIRIQIIRMEKELRAQOILEAKKKKKEEANAAXKLEAE 788  
QY 469 LEIMSEKLRRTAEDNRIYQRTQMGHONREEDNADRFPMDSIKQIHRRADKEENFEM 528  
DB 789 KRIRKEMER-----QOAVLLKH-QERERRRH---MMLMKAMEARKKAEEK--ER 833  
QY 529 LQOQERAKVVGQOQONINPSSNDCKRABEVSSFIEPOKEKEEVEEEREMLIKQOEKK 588  
DB 834 LKQEK-----DEKRLN-----KERKLEQRLELEM-AKELKRP 866  
QY 589 MEDM 592  
DB 867 NEDM 870  
  
RESULT 5  
CING\_HUMAN STANDARD; PRT; 1197 AA.  
AC Q9P2M7; Q9NR25;  
ID 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Cingulin.  
GN CGN OR KIAA1319.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MISSING (in isoform 3).  
TISSUE=Neuroepithelium;

RX MEDLINE=20499514; PubMed=11042084;  
 RA Citi S., D'Atti F., Parry D.A.D.;  
 RT "Human and Xenopus cingulin share a modular organization of the  
 RT coiled-coil rod domain: predictions for intra- and intermolecular  
 RT assembly.";  
 RL J. Struct. Biol. 131:135-145(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20181126; PubMed=10718198;  
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XVI.  
 RT The complete sequences of 150 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 7:65-73(2000).  
 RN [3]  
 RP INTERACTION WITH TUP1.  
 RX MEDLINE=22140336; PubMed=12023291;  
 RA D'Atti F., Nadalucci F., Citi S.;  
 RT "Evidence for a functional interaction between cingulin and ZO-1 in  
 RT cultured cells.";  
 RL J. Biol. Chem. 277:27757-27764(2002).  
 CC -1- FUNCTION: Probably plays a role in the formation and regulation of  
 CC the tight junction (TJ) paracellular permeability barrier.  
 CC -1- SUBUNIT: Homodimer (By similarity). Interacts with TUP1/ZO-1.  
 CC -1- TISSUE SPECIFICITY: Localized on the cytoplasmic face of tight  
 CC junctions of polarized epithelia and some endothelia. Expressed in  
 CC pancreas, kidney, liver and lung, but not in skeletal muscle,  
 CC placenta, brain or heart.  
 CC -1- DOMAIN: Deletion of the ZO-1 interaction motif (ZIM) decreases but  
 CC does not abolish colocalization with ZO-1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 CC EMBL; AF263462; AAF74498.1; ALT\_INT.  
 CC EMBL; AB037740; BAA92557.1; ALT\_INT.  
 CC Genew; HGNC:17429; CGN.  
 CC DR GO: GO:0005923; C:tight junction; NAS.  
 CC DR GO: GO:0003779; F:actin binding; IBS.  
 CC DR GO: GO:0005515; F:protein binding; IPI.  
 CC DR InterPro: IPR002928; Myosin tail.  
 CC DR Pfam; PF01576; Myosin\_tail; 1.  
 CC DR K1am; Pf01576; Myosin\_tail; 1.  
 CC FT DOMAIN 1 351 HEAD.  
 CC FT DOMAIN 352 1154 COILED COIL (POTENTIAL).  
 CC FT DOMAIN 1155 1197 TAIL.  
 CC FT DOMAIN 363 836 GLU-RICH.  
 CC FT SITE 42 56 ZIM.  
 CC FT SITE 106 400 INTERACTS WITH ZO-2.  
 CC SEQUENCE 1197 AA; 136385 MW; 0C9375283AABAFA3D CRC64;  
 SQ  
 Query Match 5.4%; Score 178; DB 1; Length 1197;  
 Best local similarity 19.0%; Pred. No. 0.046;  
 Matches 173; Conservative 121; Mismatches 284; Indels 332; Gaps 40;  
 QY  
 DB 15 QGGRPE-----VQLVQGLAGTR--LASQDQGEWEVISKKNKPKNTSGKTW 63  
 DB 31 RGRPRPADARASTGYAVRVQINGPFIYVINGEGKSGSFGVQIGANDQG-ASGA-- 87  
 QY 64 VSQNSGNPRAMGQOQGRGNSVGRGNVSGRGNGRGT---QANISRGALSKRYD 119  
 DB 88 LSSDLELPENPYQYKFPAPQSGSTSDBEAGAYWNGQLKSHSQASLAGEGVPDPSNRS 147  
 QY 120 NNFV-APPVSRP-----PLE-----GGMNQARGSGSKQHTAVQEPDVE 158  
 DB 148 NMLLELPVAVSPGSTIDTAPLVSVDLINFDSQLG---QANGRTGRTTRM--LPPEQ 202

QY 159 DDVNASEENDSDALDSDDD-----LASDDYSDVSQKSHGSRKQNKFKFPGSLDS 213  
 DB 203 RRRSKSLDSRLRDPFEERERQSTNHTWTSYTKYNHHTGTSQPAQSQN-----LSP 253  
 QY 214 LS-----TEQINPQRMWCPACQN-----GPGADMY-NLH 244  
 DB 254 LSGFSRSRQTQDWVQSFEEPRRSADPTMLQFKSTPDLNDQGEAAPPGSVDMKATY 313  
 QY 245 PLIAHARTGARRVKLHRELAIEVKDLQMRGASVIPCSE---IYQ-----WKQ 291  
 DB 314 GILREGSSESETSVR--RVSLSVLEK---WQPLVWSSGSTKAVAGQGLFRKVELEQRK 368  
 QY 292 LGEDEKQVEIWPWPVITINTRLDDKDDKVLGMQNBELLE-----Y 333  
 DB 369 LDESVKRRQLEPSQYGLRQLEKTEEC SRL---QELERRKGEAQOQNKELQNMRL 424  
 QY 334 FDKYEALR-----ARHSYGPQ----- 349  
 DB 425 LDQGEDLRHGLETVQWELQNKLGHVQGPAPAKEVLKLLLETRLELBEVLBGKQVEBQL 484  
 QY 350 --GHRGVSYL--MFESSAT-----GYLEMR-- 371  
 DB 485 RLREBELTALKALKEEVAASRDQVEHVRQYQRTQQLRRSQMDATQDHAVALERQRM 544  
 QY 372 ----LHRELAEWGLDRIAMGQRKSYFGSVQLYGFILATQDL----- 410  
 DB 545 SALVRGQRLREFTSEETGHM--QSMFQNKEDL--RATQELQLQMEKEEMEEELG 598  
 QY 411 --DIFNQH-----SOGKTR--LKEFLSYQEWVVKEL----- 438  
 DB 599 EKTVEQLRELEQAPASADDTROYAVLKKELLRRQE-ELKELQEROSQEVAGRRHREDE 657  
 QY 439 -----ROISENQO---LNYFRKLSKQNKAKYLVESLEIMSEK----- 475  
 DB 658 KQAVLAVRQADRGRLERQNLQKTQLQQLRQDCEKSKAKVAEATYVIGQRAAVET 717  
 QY 476 -LRRTRADNRIRYQRTNQHONREEM-----DAHRFFWDSQKQIHERRDAKEENFEM 528  
 DB 718 TLRQEBNDEFRRLILGLEQLKTRGLVNGGAVEARLRDKLQRLAEVQOLEALNA 777  
 QY 529 LQOQE-----RAKV-----VQGOQONINSSNDCKRKAEEVSFTFOE- 568  
 DB 778 SQEBESLSAAARALAEARLEAQRGLARLQGOEQLNLNLAEBEKKQVLRFGALBEQ 837  
 QY 569 -----KEMEFYEREMILIKQEKMEKMKRHHHEIFDLEKEP-DEALQQLM 615  
 DB 838 KRLLDRTVRLNKELEKIGEDSKQALQQLQQLLEKARKARRVADQROAKWASEAEK 897  
 QY 616 YKHGLNEDD 625  
 DB 898 TSGGLSRRLD 907  
 RESULT 6  
 MYH8 HUMAN STANDARD; PRT; 1937 AA.  
 AC P13535; Q14910; (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Myosin heavy chain, skeletal muscle, perinatal (MyHC-perinatal).  
 GN MYH8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCB1 TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=90323631; PubMed=2173371;  
 RA Karash-Mizrachi I., Peghail R., Shows T.B. Jr., Leiwand L.A.;  
 RT "Generation of a full-length human perinatal myosin heavy-chain-  
 RT encoding cDNA.";



```

RESULT 7
R50. ACUAE STANDARD; PRT; 978 AA.
ID R50. ACUAE
AC 067124;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable DNA double-strand break repair rad50 ATPase.
GN RAD50 OR AQ_1006.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxId=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.B., Overbeek R., Sneed M.A., Kellar M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -1- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -1- SIMILARITY: Belongs to the SMC family. Rad50 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000718; AAC07092.1; -.
CC PIR; A70387; A70387.
CC HAMAP; MF_00449; -.
CC DR InterPro:IPR003439; ABC transporter.
CC DR InterPro:IPR007523; DUF498.
CC DR InterPro:IPR004592; SbCC.
CC DR InterPro:IPR002017; Spectrin.
CC DR Pfam; PF04430; DUF498; 1.
CC DR TIGRFAMs; TIGR00618; sbcc; 1.
CC DR DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
CC FT NP_BIND 32 39 ATP (BY SIMILARITY).
CC FT DOMAIN 160 826 COILED COIL (POTENTIAL).
CC FT SEQUENCE 978 AA; 115897 MW; 980F2BF51ADD1151 CRC64;
SQ
Query Match 5.2%; Score 173.5; DB 1; Length 978;
Best Local Similarity 19.6%; Pred. No. 0.062;
Matches 130; Conservative 107; Mismatches 212; Indels 215; Gaps 30;
QY 106 NISGRGALSRKYD--NNFVAPPVSRPPEGGNNWQARGSSAQTAVQEPFVEDVDVN 163
DB 179 NLBGRKALKEVEYLDYV---PTKREYLE-----KTLKNLEELKEIKETEERK 225
QY 164 ASEBENDSDALDSDDDLAS-----DDYDSVVS---QKSHSRK----- 139
DB 226 LRQELKAEKESDLERELQVWTKLELNLEKEVEKREKLEFRSRVAVYPIAKIEE 285
QY 200 -----QNKWFKKFPGLSLSLSI--EQIN-----EQRQWHCPACQNGP 235
DB 286 IDKKLTLEKRAKNTLTLELAVLNDLSEFAGEELNRIRAEKEKPEKEKEKEKE----- 337
QY 236 GAIDWNLHPLLAARTGARRVYLHLELAEVLEKDIQMGASVYPCGEIYGQWKGIGED 295

```

```

DB 338 -----LEH-----RLKKLOEIKEL--KEISQLSSS-----LKEK 365
QY 296 EKDYETVMPWYIINWTRLDKDNDKWLGMGNQELLEYF-----DKYEALRARSYQ 349
DB 366 EREYQAOEPE-DLSERVEK---OKLVATEERLEKIKLFSSEETSLKMERLIVE 421
QY 350 GHRGMSVL-----MESSATGYLEAERLH-----RELAMGLDRJAWGQKRMF 393
DB 422 LQRLKELEKEGQLENLTQYKEKKVHEKVLNELKELEBELKEREHLHYAHVASYLS 481
QY 394 SGCVNQLYGFATKQDLIFQHSQCKRLKE--ELKSYQM----- 433
DB 482 PDTCPCVGGIYRGALR--NVDAGISLELHAKLEKEKEEEDITTLKLYAQKINSLE 539
QY 434 -----VKEELQISEDN-----QQLNYFKKLS--KONKAKVLE----- 467
DB 540 EMEKARNEVEBELRKIPENIKLEBELREKEKLEHLKNTKAKYKALJEDQKQKEBA 599
QY 468 -----SLEIWSKLRTRAEEDNRIVRQRTYQH---EQNREMDAHRFPMDSIKQIH 516
DB 600 KLHKAQTELELKEKIR---EKSLVKEFKELRYVERLEDEEESLKEEINYNSYLOIE 656
QY 517 ERDRAKEEFEMLQOOEAKVVGQOQONINPSNDCKRAEVSFTLEQF--KEMEEL 574
DB 657 EKEKLRKHFELSER--KSKLEGELSA--LNSINSLEERKEKLELANIYVAKSPRV 714
QY 575 VE-----EEMLIKQOEKMEEMKGRHRE-----EIFPLEKEFDEALEQL 614
DB 715 VELIYGDKEALEERKIKERESFQSLKKSIEERKLEKEBQIRLSIDIKGSEVVKQL 774
QY 615 MYKH 618
DB 775 EEKH 778

```

RESULT 8  
 MYSS RABIT STANDARD; PRT; 1084 AA.  
 ID MYSS RABIT  
 AC P02562;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Myosin heavy chain, skeletal muscle (fragments).  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxId=9986;  
 RN [1]  
 RP SEQUENCE OF 1-258.  
 RA Capony J.-P., Elzinga M.,  
 RT "The amino acid sequence of A 34,000 dalton fragment from S-2 of  
 RT myosin".  
 RL Biophys. J. 33:148A-148A(1981).  
 RN [2]  
 RP SEQUENCE OF 259-428.  
 RX MEDLINE=85131142; PubMed=3972832;  
 RA Lu R.C., Wong A.,  
 RT "The amino acid sequence and stability predictions of the hinge  
 RT region in myosin subfragment 2".  
 RL J. Biol. Chem. 260:3456-3461(1985).  
 RN [3]  
 RP SEQUENCE OF 409-1084 FROM N.A.  
 RX MEDLINE=87304245; PubMed=3305014;  
 RA Mada K., Szakiel G., Wittinhofer A.,  
 RT "Characterization of cDNA coding for the complete light meromyosin  
 RT portion of a rabbit fast skeletal muscle myosin heavy chain".  
 RL Eur. J. Biochem. 167:97-102(1987).  
 CC -1- FUNCTION: Muscle contraction.  
 CC -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2  
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)  
 CC and 2 regulatory light chain subunits (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing



RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
RA Wang Q., Wang Y., Wang Z., White J., Williamson D., Wu H., Yao Z.,  
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Mixx P.,  
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Mixx P.,  
RA Pullen R., Johnson D., Bemis D., Bentley D., Bradshaw H., Bourne S.,  
RA Cordon M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
RA Hinds K., Kamp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,  
RA Schief F., Walker C., Wamsley A., Wolddmann P., Pepin K., Nelson J.,  
RA Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,  
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Salts S.,  
RA Budarf M.L., McDonald H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
RA Edelman L., Kim U.J., Shizuya H., Simon M.I., Dunham J.P.,  
RA Peyrard M., Keda D., Serousi E., Fransson I., Tapia I., Bruder C.E.,  
RA O'Brien K.P., Wilkinson P., Bodewicht A., Hartman K., Hu X.,  
RA Khan A.S., Lane L., Tjallum Y., Wright H.,  
RT "The DNA sequence of human chromosome 22." ;  
RL Nature 402:489-495(1999).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Colon; PubMed=12477932;  
RX MEDLINE=22386257; PubMed=12477932;  
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Jordan B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jorde H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marzina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stachenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.J.,  
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Helton E., Kettelman M., Medan A., Rodriguez Y., Bouffard G.G.,  
RA Whiting M., Medan A., Young A.C., Shevchenko I., Bouffard G.G.,  
RA Blakesley A.C., Grimwood J.W., Schmutz J., Myers R.M.,  
RA Rodriguez A.C., Krzywinski M.I., Skalska U., Smalls D.E.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
RA Scheraga A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [6]  
RP SEQUENCE OF 340-563 FROM N.A., AND SUBCELLULAR LOCATION.  
RX TISSUE=Fetal brain;  
RX MEDLINE=99068504; PubMed=9853615;  
RA Ueki N., Oda T., Kondo M., Yano K., Noguchi T., Muramatsu M.A.,  
RT "selection system for genes encoding nuclear-targeted proteins." ;  
RL Nat. Biotechnol. 16:138-1342(1998).  
CC -1- FUNCTION: May regulate actin cytoskeletal organization, cell  
CC spreading and cell contraction by directly binding and stabilizing  
CC filamentous F-actin. The localized formation of TARA and TRIO  
CC complexes coordinates the amount of F-actin present in stress  
CC fibers. May also serve as a linker protein to recruit proteins  
CC required for F-actin formation and turnover.  
CC -1- SUBUNIT: Binds to TRIO and F-actin. May also interact with myosin  
CC II.  
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic, localized to F-  
CC actin in a periodic pattern.  
CC TISSUE SPECIFICITY: Widely expressed. Highly expressed in heart  
CC and placenta.  
CC -1- DOMAIN: Contains at least 2 actin-binding sites per coiled-coil  
CC dimer.  
CC -1- MISCELLANEOUS: Has been identified in Ref.6 by a selection system  
CC for genes encoding nuclear-targeted protein.  
CC -1- SIMILARITY: Contains 1 PH domain.  
CC -1- CAUTION: Ref.6 sequence differs from that shown due to a  
CC frameshift in position 478.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

[illegible]



Db 550 QDVYVELSHIKTRSERIEQLKEHLRLAMALQEKSMRN 589

## RESULT 10

TRHY\_RABIT STANDARD; PRT; 1407 AA.

AC P37709;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Trichohyalin.  
 GN THH.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Fietz M.J., Rogers G.B.;  
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Intermediate filament-associated protein that associates  
 in regular arrays with keratin intermediate filaments (KIF) of the  
 inner root sheath cells of the hair follicle and the granular  
 layer of the epidermis. It later becomes cross-linked to KIF by  
 isodipeptide bonds. It may serve as scaffold protein, together  
 with involucrin, in the organization of the cell envelope or even  
 anchor the cell envelope to the KIF network. It may be involved in  
 its own calcium-dependent postsynthetic processing during terminal  
 differentiation.  
 CC -1- SUBUNIT: Homodimer (Probable).  
 CC -1- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as  
 the inner root sheath (IRS) of hair follicles and medulla, and in  
 the filiform papillae of dorsal tongue epithelium (Probable).  
 CC -1- DEVELOPMENTAL STAGE: Expressed during late differentiation of  
 the epidermis.  
 CC -1- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand  
 calcium-binding domains. Domains 2-4, 6, and 8 are almost entirely  
 alpha-helical, configured as a series of peptide repeats of  
 varying regularity, and are thought to form a single-stranded  
 alpha-helical rod stabilized by ionic interactions. Domain 6 is  
 the most regular and may bind KIF directly by ionic interactions.  
 CC Domains 5 and 7 are less well organized and may induce folds in  
 the molecule. Domain 9 contains the C-terminus, conserved among  
 different species.  
 CC -1- PTM: Substrate of transglutaminase. Some 200 arginines are  
 probably converted to citrullines by peptidylarginine deiminase.  
 CC -1- SIMILARITY: In the N-terminal section; belongs to the S-100  
 family.  
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z19092; CAA79519.1; -.  
 CC DR PIR; S28589; S28589.  
 CC DR HSSP; P02633; 4ICB.  
 CC DR InterPro; IPR001751; GABP\_S100.  
 CC DR InterPro; IPR002048; EF-hand.  
 CC DR Pfam; PF00036; ehand; 1.  
 CC DR Pfam; PF01023; S\_100; 1.  
 CC DR ProDom; PD003407; GABP\_S100; 1.  
 CC DR PROSITE; PS00018; EF\_HAND; 1.  
 CC DR PROSITE; PS00303; S100\_GABP; 1.  
 CC KW Keratinization; Calcium-binding; Repeat; Citrullination.  
 CC FT DOMAIN 1 91 S-100 LIKE.  
 CC FT CA\_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).  
 CC FT CA\_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).

SQ SEQUENCE 1407 AA; 183781 MW; AE17D2A159F12B7F CRC64;

Query Match 5.2%; Score 172.5; DB 1; Length 1407;  
 Best Local Similarity 21.0%; Pred. No 0.1; Indels 141; Gaps 17;  
 Matches 87; Conservative 71; Mismatches 116;

QY 278 SVIPCEIYGQW-----KGIQEDKDYELVW-----PPVVIIMNTRLDKDN 319  
 Db 7 SIIDIIEIFNQYASHDCDAVLKKDKILDLBREGAVLQRPHPDPTVDMLELDR-DS 65  
 QY 320 DKMLGKNOBLELPFKYELARRHSYGPQGRGMSVLPFSSATYILEARHLRLAEM 379  
 Db 66 DGLVG-----PDEF-----CLIFKLAQAAYVALGQ-----AS 93  
 QY 380 GLDRIIMGKRSMFSGVAVLYGFLATKQDLDIFNHSQGRRL-----KPEIKS 429  
 Db 94 GLDE---EKRS-----HSGKGLLONRQEDQRRLELD 125  
 QY 430 YQ-----EMVVELNQISENQOLNFKNLSKO-NKRAKYLESLF-----470  
 Db 126 RQPEDEPERRRMQKQGERELAEERQKREFEQHYGRQYRDKQRLQROLEERRAE 185  
 QY 471 -----IMSEKLRTAEDNRIVRQRTMQEKNREEMDAIDRFMSIKQI 515  
 Db 186 EQLRRRKGRLAEFLIEBQLRR--NEQELKRELIEEQRRERREOHERALQEEERQI 243  
 QY 516 HERRDAKENFEMLQOQERBAKVVGQOQONINSNDCKRRAEVSFLIEFQF---KEM 571  
 Db 244 LRQRNRREEREOQLRLRELEIRERERQLEOEERREOQLREORERREOQLRRL 303  
 QY 572 EEFVEEREMLIK-----DOEKMEEMKRRHHEIFDLK--EFDALAEQLM 615  
 Db 304 EE-IRREERQLEBERREERQLEBERREERQLEBERREERQLEBERREERQLE 357  
 RESULT 11  
 MY52 DICDI STANDARD; PRT; 2116 AA.  
 ID MY52 DICDI  
 AC P08759;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Myosin II heavy chain, non muscle.  
 DE MHCA.  
 GN Dictyostelium discoideum (Slime mold).  
 OS Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.  
 OC NCBI\_TaxID=44689;  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87092266; PubMed=3540939;  
 RA Warrick H.M., de Lozanne A., Leitwand L.A., Spudich J.A.;  
 RT "Conserved protein domains in a myosin heavy chain gene from  
 Dictyostelium discoideum.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).  
 RN [2]  
 RP PHOSPHORYLATION SITES, AND MUTAGENESIS.  
 RC STRAIN=AX2.  
 RX MEDLINE=90353583; PubMed=2387408;  
 RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,  
 RA Gerisch G.;  
 RT "Replacement of threonine residues by serine and alanine in a  
 phosphorylatable heavy chain fragment of Dictyostelium myosin II.";  
 RL FEBS Lett. 269:239-243 (1990).  
 RN [3]  
 RP PHOSPHORYLATION SITES.  
 RX MEDLINE=88112226; PubMed=2828113;  
 RA Wagle G., Noegel A., Scheel J., Gerisch G.;  
 RT "Phosphorylation of threonine residues on cloned fragments of the  
 Dictyostelium myosin heavy chain.";  
 RL FEBS Lett. 227:71-75(1988).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.  
 RX MEDLINE=95345066; PubMed=7619795;



RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Suton K., Holden H.M.,  
 RA Rayment I.; Rayment I.; Rayment I.; Rayment I.; Rayment I.; Rayment I.;  
 RT "X-ray structures of the myosin motor domain of Dictyostelium  
 RT discoidium complexed with MgADP.Befx and MgADP.ALf4-";  
 RL Biochemistry 34:8960-8972 (1995).  
 RN [5]  
 RN X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.  
 RP MEDLINE=95345067; PubMed=7619796;  
 RX Smith C.A., Rayment I.;  
 RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the  
 RT truncated head of Dictyostelium discoidium myosin to 2.7-A  
 RT resolution.";  
 RL Biochemistry 34:8973-8981 (1995).  
 RN [6]  
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.  
 RX Smith C.A., Rayment I.;  
 RT "X-ray structure of the magnesium(II)-ADP.vanadate complex of the  
 RT Dictyostelium discoidium myosin motor domain to 1.9-A resolution.";  
 RL Biochemistry 35:5404-5417 (1996).  
 RN [7]  
 RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.  
 RP MEDLINE=97452580; PubMed=9305951;  
 RX Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;  
 RT "X-ray structures of the MgADP, MgATPgammass, and MgAMPNP complexes  
 RT of the Dictyostelium discoidium myosin motor domain.";  
 RL Biochemistry 36:11619-11628 (1997).  
 RN [8]  
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.  
 RP MEDLINE=98070605; PubMed=9405148;  
 RX Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;  
 RT "X-ray crystal structure and solution fluorescence characterization  
 RT of Mg.2'(3)-O-(N-methylanthraniloyl) nucleotides bound to the  
 RT Dictyostelium discoidium myosin motor domain.";  
 RL J. Mol. Biol. 274:394-407 (1997).  
 CC -1- FUNCTION: Myosin is a protein that binds to actin and has ATPase  
 CC activity that is activated by actin.  
 CC -1- SUBUNIT: Myosin II heavy chain is two-headed. It self-assembles  
 CC into filaments. Hexamer of 2 heavy chain subunits (MHC), 2 alkali  
 CC light chain subunits (MLC) and 2 regulatory light chain subunits  
 CC (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL  
 CC CORTEX.  
 CC -1- DOMAIN: Each myosin heavy chain can be split into 1 light  
 CC meromyosin (LMM) and 1 heavy meromyosin (HMM). It can be further  
 CC split into 2 globular subfragments (S1) and 1 rod-shaped  
 CC subfragment (S2).  
 CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing  
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,  
 CC characteristic for alpha-helical coiled coils.  
 CC -1- PTM: Phosphorylation inhibits thick filament formation and reduces  
 CC the actin-activated ATPase activity.  
 CC -1- MISCELLANEOUS: Dictyostelium myosin II has no K(2)EDTA ATPase  
 CC activity, perhaps correlated with the absence of a Cys at the SH-1  
 CC position (688).  
 CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.  
 CC -1- SIMILARITY: Contains 1 IQ domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M14628; AAA3227.1; -.  
 DR PIR: A26655; A26655.  
 DR PDB: 1MMA; 03-DEC-97.  
 DR PDB: 1MMD; 17-AUG-96.  
 DR PDB: 1MNG; 03-DEC-97.  
 DR PDB: 1MNN; 03-DEC-97.  
 DR PDB: 1MND; 17-AUG-96.

DR PDB: 1MNE; 17-AUG-96.  
 DR PDB: 1VOM; 23-DEC-96.  
 DR PDB: 1LYK; 28-JAN-98.  
 DR PDB: 1DOX; 20-DEC-00.  
 DR PDB: 1DOY; 20-DEC-00.  
 DR PDB: 1DOZ; 20-DEC-00.  
 DR PDB: 1DIA; 20-DEC-00.  
 DR PDB: 1DIB; 20-DEC-00.  
 DR PDB: 1DIC; 20-DEC-00.  
 DR PDB: 1FMV; 20-DEC-00.  
 DR PDB: 1FMW; 20-DEC-00.  
 DR PDB: 1G8X; 17-JAN-01.  
 DR PDB: 1UWY; 07-NOV-01.  
 DR PDB: 1UX2; 07-NOV-01.  
 DR DictyBase; DDB0002015; mbca.  
 DR InterPro; IPR000048; IQ\_region.  
 DR InterPro; IPR001609; myosin\_head.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR008989; Myosin\_S1\_N.  
 DR Pfam; PF00612; IQ; 2.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF02736; Myosin\_N; 1.  
 DR PRINTS; PR00193; MYOSTINHEAVY.  
 DR PRODOM; PD000355; myosin\_head; 1.  
 DR SMART; SM00242; MYSC; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR PROSITE; PS0096; IQ; 1.  
 KW Myosin, coiled coil, Actin-binding, 3D-structure;  
 KW Calmodulin-binding, Methylation, Phosphorylation,  
 KW MYOSIN HEAD-LIKE.  
 FT 1  
 FT 761  
 FT 762  
 FT 791  
 FT 2116  
 FT 186  
 FT 638  
 FT 660  
 FT 738  
 FT 752  
 FT 130  
 FT 130  
 FT 1823  
 FT 1823  
 FT 1833  
 FT 2029  
 FT 3  
 FT 7  
 FT 8  
 FT 10  
 FT 15  
 FT 28  
 FT 37  
 FT 42  
 FT 41  
 FT 55  
 FT 59  
 FT 63  
 FT 73  
 FT 74  
 FT 76  
 FT 78  
 FT 83  
 FT 85  
 FT 87  
 FT 90  
 FT 93  
 FT 99  
 FT 110  
 FT 111  
 FT 113  
 FT 116  
 FT 119  
 FT 120  
 FT 121  
 FT 126  
 FT 132  
 FT 137  
 FT 142  
 FT 143  
 FT 145  
 FT 148  
 FT 150  
 FT 155  
 FT 169  
 FT 170  
 FT 179  
 FT 181  
 FT 182  
 FT 185  
 FT 200  
 FT 206  
 FT 226  
 FT 228  
 FT 234  
 FT 235  
 FT 236  
 FT 240  
 FT 247  
 FT 249  
 FT 250

COILED COIL (POTENTIAL).  
 ATP.  
 ACTIN-BINDING.  
 ACTIN-BINDING.  
 METHYLATION (DI-) (POTENTIAL).  
 METHYLATION (BY MHCK).  
 PHOSPHORYLATION (BY MHCK).  
 PHOSPHORYLATION (BY MHCK).  
 PHOSPHORYLATION (BY MHCK).

```

FT STRAND 253 261
FT HELIX 265 268
FT TURN 269 269
FT TURN 273 274
FT STRAND 278 278
FT HELIX 279 287
FT HELIX 290 296
FT HELIX 297 297
FT TURN 305 303
FT TURN 316 317
FT HELIX 320 334
FT TURN 335 335
FT HELIX 338 355
FT TURN 356 356
FT STRAND 360 360
FT STRAND 368 368
FT HELIX 373 382

```

Query Match 5.2%; Score 171.5; DB 1; Length 2116;  
 Best Local Similarity 20.3%; Pred. No. 0.19; Indels 173; Gaps 25;  
 Matches 114; Conservative 101; Mismatches 173;

```

QY 156 DVEDVDVNASEENDSDALDDSDDLAS--DD----YDSVDVQSHSGSRKQNFKKFG 209
DB 983 EVELSEFSESESKDGVLEKTRVRLQSELDLTVRLDSTKXKSELRLQKLEEE--- 1039
QY 210 SLDSLSIEQINEQOROWHCPACQNGRGAIDWYMLHPLAARTKCARV---KLHRLAE 266
DB 1040 -----LKQVQ-----ALAEATAKLQEAANKKLQGEYTE 1070
QY 267 VLEK---DLQMRASVYPGGEIYGQWKGLEDKDYETVWPVYIIMNTRLDKDDDKML 323
DB 1071 LNEKFSSEVAT-----SNVKSKTLE---SGLVAVNELDESK--- 1108
QY 324 GNGNOELLEYFDYKYLRLAR-----HSYGPQGHGMSVLMFESSATGYLAEHLRE 375
DB 1109 ---NRDALB---KKKKLLDNLWLEMKQDLESTG---GEKSLYDLKVGQESDWEALRNQ 1158
QY 376 LAAM-----GLDRIANGQKSMFSGVRYQYFLATQDDIDINQSHQKTRIKFELKSY 430
DB 1159 ISELQSTIANLEKTI-----KSTLEGEVARIQGELEAEQLAKSNVVEKQKKVEILDLEDKSA 1213
QY 431 Q-----EMVVKELRQ-ISEDNQOLNYPKNTLKSQNKAKVLEES-----LEIMS 473
DB 1214 QLAEEFPAKQALDKLKKQLQELSEVOTQJSEANNKVNDSSTNKLHETSFPNNLKELEA 1273
QY 474 EKLRTAEDNRIVRORTKQH-----EQNREMDAHDHFPMDSIKQIHERRD----- 520
DB 1274 EQAKAQALEKKRGLGSELKGVHNEQLEEEKKESNEKRVLDLEKVESLKQIEEVAS 1333
QY 521 -----AKENFEMLQOQERAKVYGGQOONIN-----PSSNDCKRAEYVSSTIEF 566
DB 1334 KKAATKAKNKESELLEIKRQ-YADVYSRDKSVEQLKTLQALQNEELRNFAEAEQGLDR 1392
QY 567 QEK-----EMSEFV-----EEREMLIKQ-----EKMEEDK----- 593
DB 1393 AERSKKAEFDLEAAVNLLEETAKKAKKAKKAETDYRTKSELDLDAKNVSSQYVQ 1452
QY 594 -KRNHEIFDLKEKFDALQEQ 613
DB 1453 IKRLNELSELRLVLEADER 1473

```

RESULT 12  
 MYH4 RABIT  
 ID MYH4 RABIT STANDARD; PRT; 1938 AA.  
 AC 028641;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Myosin heavy chain, skeletal muscle, juvenile.  
 OS *Oryctolagus cuniculus* (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 NCBI\_TaxId=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white; TISSUE=skeletal muscle;  
 RA Maeda K., Hostinova E., Roesch-Kleinhauf A., Schuster H., Gaasperik J.,  
 RA Wittunghofer A.;  
 RT Isolation, sequencing of myosin heavy chain cDNA from rabbit  
 RT skeletal muscle and a novel cosynthesis of S-1 fragment with the  
 RT essential and regulatory light chains.";  
 RL Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.  
 CC 1- FUNCTION: Muscle contraction.  
 CC 1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2  
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)  
 CC and 2 regulatory light chain subunits (MLC-2).  
 CC 1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC 1- DOMAIN: The rodlike tail sequence is highly repetitive, showing  
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,  
 CC characteristic for alpha-helical coiled coils.  
 CC 1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light  
 CC meromyosin (LM) and 1 heavy meromyosin (HMM). It can later be  
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped  
 CC subfragment (S2).  
 CC 1- SIMILARITY: Contains 1 myosin-like globular head domain.  
 CC 1- SIMILARITY: Contains 1 IQ domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; U32574; AA074199.1; -  
 DR PIR; A59293; A59293.  
 DR HSSP; P1358; 2MYS.  
 DR InterPro; IPR000048; IQ\_region.  
 DR InterPro; IPR001609; myosin\_head.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR002928; Myosin\_tail.  
 DR Pfam; PF00612; IQ\_2.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF02736; Myosin\_N; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR ProDom; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00242; MYSC; 1.  
 DR PROSITE; PS50096; IQ; 1.  
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.  
 FT DOMAIN 1 783  
 FT MYOSIN HEAD-LIKE.  
 FT DOMAIN 784 813  
 FT COILED COIL (POTENTIAL).  
 FT NP\_BIND 842 1938  
 FT ATP (POTENTIAL).  
 FT DOMAIN 179 186  
 FT ACTIN-BINDING (BY SIMILARITY).  
 FT DOMAIN 658 680  
 FT ACTIN-BINDING (BY SIMILARITY).  
 FT MOD\_RES 760 774  
 FT METHYLATION (MONO-) (BY SIMILARITY).  
 FT MOD\_RES 35 35  
 FT METHYLATION (TRIM-) (BY SIMILARITY).  
 FT MOD\_RES 130 130  
 FT METHYLATION (TRIM-) (BY SIMILARITY).  
 FT MOD\_RES 552 552  
 FT METHYLATION (TRIM-) (BY SIMILARITY).  
 FT MOD\_RES 756 756  
 FT METHYLATION (TRIM-) (BY SIMILARITY).  
 SQ  
 SEQUENCE 1938 AA; 223064 MW; DBA8A2EC5B182626 CRC64;

Query Match 5.1%; Score 169; DB 1; Length 1938;  
 Best Local Similarity 21.1%; Pred. No. 0.23;  
 Matches 121; Conservative 94; Mismatches 183; Indels 175; Gaps 28;

```

QY 152 QEPVEDVDVNASEEND-----SDALDDSDVDSVQSHSGSRKQNF-- 202
DB 1309 QAFTOOIEELKQLEIEIRAKSALAHVQLQSAHSDCLRLQEQYEBQEAQVLEQRAWMSKAN 1368

```

```

QY 203 -----WFKKFGSLDLSIQINEPQRQWCPACQNGPAGAIWNLHPLHAKTKGARR 257
DQ 1369 SEVAQMRTRYK-----ETDAIQRTSE-----LEBAKKAQOR 1399
QY 258 VKIARELAETLENDLQMGASVYPOGEIYQMGKLGDEKQYIWPVWVIMNTR----- 313
DQ 1400 LQDAEHEVEAVN-----AKCSLEKTKORLQNEVDL-----MIDVERINACA 1443
QY 314 -----LQKDNNDKWLGMGNOLLEYPFKYALPARHSYQGRGMSVLMF-----ESS 362
DQ 1444 ALKKQRN-----FDKILAEWKHKYETTHAEASQKESRSISTEVEFKYKNAYESLDQL 1498
QY 363 ATGYLEAERLARELAENGELDRIMGQR-----SMFGGVRQLYGFL-ATKQDLIDFNQHS 417
DQ 1499 ETIKREKKNLQGEISDL-TEQIAEGGRIHELEKVKQVQEKSELOALAEASLEHE 1557
QY 418 QGR-TRIKPELKYQGVENVKELQISENQOLNYPFKNLSKONKAKVLEESLEISKL 476
DQ 1558 EGKILRIQLELNOVKSEIDKKIAKDELDL-----KRI-HIRVESHQSTLDAL 1608
QY 477 RRTAEDRIIVROPTKQHQEQNREEMAH--DRPFMSIK-----QIH----- 516
DQ 1609 R--SRNDARIRKKQMGDLNEMEIQLNHNAMAAEALRYVNTQGLKDTQLHLDALR 1665
QY 517 -----ERR-----D-K-ENFEMIQOQERAVVQOQ-----ONIN 546
DQ 1666 QGQDLKEQLAMVERANVILQAEIHELRLATLEQTERSKVAEQELDASERVQLHTON-- 1723
QY 547 PSSNDCKRAEVEVSFIEQEKEMEEFEE-----REMLIKQEKK 588
DQ 1724 -TSLINKKKKLETTISQI--QGMEDIVQBARAAEKAKKAITDAMMAEELKKEQDTS 1779
QY 589 --MEDMKRRHHEIFDLKEFDEALQOLMYNG 619
DQ 1780 AHLERKKMGQYVLDQHRIDEA-EQALAKG 1811

```

```

CC anchor the cell envelope to the KIF network. It may be involved in
CC its own calcium-dependent postsynthetic processing during terminal
CC differentiation. (Probable).
CC -1- SUBUNIT: Monomer. (Probable).
CC -1- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as
CC the inner root sheath (IRS) of hair follicles and medulla, and in
CC the filiform papillae of dorsal tongue epithelium (Probable).
CC -1- DEVELOPMENTAL STAGE: Expressed during late differentiation of
CC the epidermis.
CC -1- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand
CC calcium-binding domains. Domains 2-4, 6, and 8 are almost
CC entirely alpha-helical, configured as a series of peptide repeats
CC of varying regularity, and are thought to form a single-stranded
CC alpha-helical rod stabilized by ionic interactions. Domain 6 is
CC the most regular and may bind KIF directly by ionic interactions.
CC Domains 5 and 7 are less well organized and may induce folds in
CC the molecule. Domain 9 contains the C-terminus, conserved among
CC the molecule.
CC -1- PTM: Substrate of transglutaminase. Some 200 arginines are
CC probably converted to citrullines by peptidylarginine deiminase.
CC -1- SIMILARITY: In the N-terminal section; belongs to the S-100
CC family.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; L09190; AAA65582.1; -.
DR PIR; A45973; A45973.
DR HSSP; P02633; 4ICB.
DR GENE; HGNC:11791; THH.
DR MIM; 190370; -.
DR GO; GO:0005856; C:cytoskeleton; NAS.
DR GO; GO:0005509; F:calcium ion binding; TAS.
DR InterPro; IPR001751; CABP_S100.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00036; ehfhand; 1.
DR Pfam; PF01023; S_100; 1.
DR ProDom; PD003407; CABP_S100; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS00303; S100_CABP; 1.
KM Keratinization; Calcium-binding. Repeat; Citrullination.
FT DOMAIN 1 91
FT CA_BIND 22 33
FT CA_BIND 62 73
FT DOMAIN 314 390
FT REPEAT 314 326
FT REPEAT 327 339
FT REPEAT 340 351
FT REPEAT 352 364
FT REPEAT 365 377
FT REPEAT 378 390
FT REPEAT 391 444
FT REPEAT 391 396
FT REPEAT 397 402
FT REPEAT 403 408
FT REPEAT 409 414
FT REPEAT 415 420
FT REPEAT 421 426
FT REPEAT 427 432
FT REPEAT 433 438
FT REPEAT 439 444
FT DOMAIN 444 702
FT DOMAIN 923 1162
FT REPEAT 923 952
FT REPEAT 953 982

```

```

FT REPEAT 983 1012 4-3.
FT REPEAT 1013 1042 4-4.
FT REPEAT 1043 1072 4-5.
FT REPEAT 1073 1102 4-6.
FT REPEAT 1103 1132 4-7.
FT REPEAT 1133 1162 4-8.
FT DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT CONFLICT 1752 1752 F -> L (IN REF. 2).
FT CONFLICT 1794 1801 QERDROYR -> RSETSTG (IN REF. 2).
FT CONFLICT 1857 1857 O -> K (IN REF. 2).
FT CONFLICT 1880 1880 V -> G (IN REF. 2).
SQ SEQUENCE 1898 AA; 247219 MW; A74B5947FB62B31D CRC64;

Query Match 5.1%; Score 167.5; DB 1; Length 1898;
Best Local Similarity 20.9%; Pred. No. 0.26;
Matches 98; Conservative 85; Mismatches 162; Indels 123; Gaps 22;

QY 246 LLAHARTGKARVYKHELAELVEKDLQ-----NRGASVIPCCEIYQWKGLEDEKDYE 300
DB 1140 LLREERK-RRQELERQYRE--EELQQRKQRYDEDDQSDLKQW-----EPEKENA 1192
QY 301 IWPMYIIMTRLDKNDKMLGMQELLEYDKTEALR-----ARSTYGPQHNG-- 353
DB 1193 V-----RNNKYCKGRENEQFQLEDQVDRDQSQDQLQHLGEGQERDRE 1238
QY 354 MSYLFESSATGYLAERLHELAEMGLDRIAMQ-----KSNMFGGVQ 399
DB 1239 QERRRMOQANRHPPEEQLEREBEKAKRDRKSQEEKQLREBEREKRRQETDRKFRE 1298
QY 400 LYGLATKODLIFNQ-----HSQKTRLKE-----LKSQYQEMV 435
DB 1299 EEOQLQREERQPLLRQDRKFRREELHQQGKFLPEEQRLREBERKFLKEEQQLRL 1358
QY 436 KEARQISED-----NOOLN-----YFNKLSKQNKAKVLEESLEIMSEQLR 478
DB 1359 EEEBQQLQDRKFRREEQQLSRQERDRKFRREEQQVRQREERKFLPEEQQLQREHRK 1418
QY 479 TAEENRIVRQTKWQ-HEQNRREMDADRFPMDSIKOI-HEBRDAK-----EEN 525
DB 1419 FRREBQLQREERQQLHRQER-----DRKFLPEEQQLRQERDRKREBQLMSQEBRK 1472
QY 526 F---EMLQOQERAKVVGQOQINPSSNDCRKAQVSSFLFEQKEMEFPVEEREML 581
DB 1473 FLREBQQLHRQQRKFLPEEQQLRRQERQ-QRQQRDRKFRRE--EQQLQREBQQLS 1529
QY 582 IKQOEK--MEDMKRHEELFDLEKFFDALBQIMTKG---LHND 624
DB 1530 ROERDRKFRLEQKVRQEQ---ERKMEDEQQLRQEQOQLROED 1573

RESULT 14
MYHD HUMAN STANDARD; PRT; 1938 AA.
AC Q9UQK3; O95252;
DT 16-OCT-2001 (Ref. 40, Created)
DT 16-OCT-2001 (Ref. 40, Last sequence update)
DT 15-MAR-2004 (Ref. 43, Last annotation update)
DE Myosin heavy chain, skeletal muscle, extraocular (MYHC-ec).
GN MYH13.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Extraocular muscle;
RX MEDLINE=99318869; Pubmed=10388558;
RA Weiss A., Schiaffino S., Leinwand L.A.;
RT "Comparative sequence analysis of the complete human sarcomeric myosin
heavy chain family: implications for functional diversity.";
RL J. Mol. Biol. 290:61-75(1999).
[2]
SEQUENCE OF 1917-1938 FROM N.A.

```

```

RC TISSUE=Extraocular muscle;
RX MEDLINE=99026150; Pubmed=9806854;
RA Winters L.M., Briggs M.M., Schachar F.;
RT "The human extraocular muscle myosin heavy chain gene (MYH13) maps to
the cluster of fast and developmental myosin genes on chromosome 17.";
RL Genomics 54:188-189(1998).
CC -1- FUNCTION: Muscle contraction.
CC -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
and 2 regulatory light chain subunits (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing
cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
characteristic for alpha-helical coiled coils.
CC -1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be
split further into 2 globular subfragments (S1) and 1 rod-shaped
subfragment (S2).
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF111782; AAD29948.1; -.
DR EMBL; AF075248; AAC83241.1; -.
DR HSSP; P13538; 2MYS.
DR Genew; HGNC:7571; MYH13.
DR MIM; 603487; -.
DR GO; GO:0005859; C:muscle myosin; TAS.
DR GO; GO:0003779; F:actin binding; TAS.
DR GO; GO:0005524; F:ATP binding; NAS.
DR GO; GO:0005516; F:calmodulin binding; NAS.
DR GO; GO:0003776; F:muscle motor activity; TAS.
DR GO; GO:0006936; F:muscle contraction; TAS.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; myosin_N.
DR InterPro; IPR002928; myosin_tail.
DR Pfam; PF00612; IQ_2.
DR Pfam; PF00663; myosin_head; 1.
DR Pfam; PF02736; myosin_N; 1.
DR Pfam; PF01576; myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR Prodom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ_1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IQ_1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Multigene family.
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
FT DOMAIN 785 814 COILED COIL (POTENTIAL).
FT NP BIND 843 1938 ATP (POTENTIAL).
FT DOMAIN 179 186 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 659 681 ACTIN-BINDING (BY SIMILARITY).
FT MOD RES 761 775 METHYLATION (TR1-) (POTENTIAL).
FT MOD RES 130 130 METHYLATION (TR1-) (POTENTIAL).
SQ SEQUENCE 1938 AA; 223678 MW; 1F6D006416381C05 CRC64;

Query Match 5.0%; Score 166.5; DB 1; Length 1938;
Best Local Similarity 19.1%; Pred. No. 0.31;
Matches 118; Conservative 98; Mismatches 204; Indels 199; Gaps 26;

QY 115 SRKYD-----NFFVAPPVSRPPLGGMNQARGSSAHTVAQEPDVEDVDNASEEB 168
DB 552 NKLYDQHLKSNFQKPK-----AKGAEAHFSLVHAGTVD-----Y 590
QY 169 NDSDALDSDDDLAS-----DDYDSVDSQKSHGRKQNKWFKKFGS 210

```



[illegible]

Search completed: April 6, 2004, 19:33:55  
Job time : 22 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: April 6, 2004, 19:30:55 ; Search time 45 Seconds  
(without alignments)  
4382.197 Million cell updates/sec

Title: US-10-030-829-3

Sequence: 1 MESSRAGPMSEKXNVQGGYRP.....EFDEALQMLYKHLNEDD 625

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3313	100.0	625	10 Q9LXD1	Q9LXD1 arabidopsis
2	778	23.5	304	10 Q7XY17	Q7XY17 triticum ae
3	548.5	16.6	170	10 Q7XYE6	Q7XYE6 triticum ae
4	242	7.3	647	10 Q8VZ79	Q8VZ79 arabidopsis
5	233.5	7.0	1036	10 Q9SMM2	Q9SMM2 arabidopsis
6	215.5	6.5	629	10 Q9SMM2	Q9SMM2 arabidopsis
7	211	6.4	634	10 Q9SMM2	Q9SMM2 arabidopsis
8	208	6.3	1003	12 Q9LXD1	Q9LXD1 arabidopsis
9	203	6.1	628	10 Q8LSD3	Q8LSD3 kapost's sa
10	198	6.0	1036	12 Q9DUM3	Q9DUM3 zea mays (m
11	194.5	5.9	447	10 Q8HBB2	Q8HBB2 kapost's sa
12	192	5.8	634	10 Q9SMM2	Q9SMM2 arabidopsis
13	192	5.8	2760	5 Q8LSD3	Q8LSD3 kapost's sa
14	190	5.7	635	10 Q9C7B0	Q9C7B0 arabidopsis
15	190	5.7	639	10 Q8W563	Q8W563 triticum mo
16	189.5	5.7	638	10 Q9LHB1	Q9LHB1 arabidopsis

17	187	5.6	694	10 Q9AX36	Q9AX36 oryza sativ
18	186.5	5.6	554	10 Q9M150	Q9M150 arabidopsis
19	186	5.6	761	13 P79793	P79793 gallus gall
20	186	5.6	1941	13 Q8UW40	Q8UW40 gallus gall
21	183.5	5.5	561	10 Q9LTD0	Q9LTD0 arabidopsis
22	183	5.5	559	10 Q04614	Q04614 arabidopsis
23	181	5.5	976	12 Q9DUN0	Q9DUN0 kapost's sa
24	180.5	5.4	895	10 Q48878	Q48878 sorghum bic
25	180	5.4	662	10 Q23064	Q23064 arabidopsis
26	179.5	5.4	188	6 Q28021	Q28021 bos taurus
27	179	5.4	1777	5 Q8LSD3	Q8LSD3 kapost's sa
28	178.5	5.4	1957	5 Q04010	Q04010 onchocerca
29	178	5.4	985	5 Q9U0S5	Q9U0S5 mytilus gal
30	178	5.4	1705	5 Q9U0S7	Q9U0S7 mytilus gal
31	177	5.3	915	5 Q869R0	Q869R0 dictyostel
32	177	5.3	1129	12 Q9GR71	Q9GR71 kapost's sa
33	176	5.3	1089	12 Q40947	Q40947 kapost's sa
34	176	5.3	1464	5 Q8LIF6	Q8LIF6 plasmodium
35	175.5	5.3	1503	5 Q861Q0	Q861Q0 dictyostel
36	174	5.3	1375	13 Q9QY37	Q9QY37 brachydanio
37	174	5.3	1939	6 Q9TV63	Q9TV63 sus scrofa
38	173	5.2	488	5 Q45420	Q45420 caenorhabdi
39	172.5	5.2	607	11 Q8BS10	Q8BS10 mus musculu
40	172.5	5.2	1156	16 Q68878	Q68878 aquilex aeo
41	172.5	5.2	1871	10 Q9SRD5	Q9SRD5 arabidopsis
42	172	5.2	892	11 Q63939	Q63939 rattus sp.
43	171.5	5.2	1391	11 Q922J3	Q922J3 mus musculu
44	171.5	5.2	1939	6 Q9TV61	Q9TV61 sus scrofa
45	171.5	5.2	1946	5 Q97291	Q97291 plasmodium

## ALIGNMENTS

RESULT 1	ID	Q9LXD1	PRELIMINARY;	PRT;	625 AA.
AC	Q9LXD1	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DE	01-JUN-2003 (TREMBLrel. 24, Last annotation update)				
EM	EMBL CAB62356.1 (SGS3) (Hypothetical protein).				
GN	AT5G233570.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsi.				
OX	NCBI_TaxID=3702;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=COLUMBIA;				
RX	MEDLINE=20181125; PubMed=10718197;				
RA	Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,				
RA	Tabata S.;				
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence				
RT	features of the regions of 3,076,755 bp covered by sixty Pl and TAC				
RT	clones.";				
RL	DNA Res. 7:31-63(2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Columbia;				
RX	MEDLINE=20306668; PubMed=10850495;				
RA	Mourrain P., Becil C., Elmayan T., Feuerbach F., Godon C.,				
RA	Morel J.-B., Jonette D., Lacombe A.-M., Nikic S., Picault N.,				
RT	Remoue K., Santal M., Vo T.-A., Vaucheret H.;				
RT	"Arabidopsis SGS2 and SGS3 genes are Required for Posttranscriptional				
RL	Gene Silencing and Natural Virus Resistance.";				
RN	Cell 101:533-542(2000).				
RP	SEQUENCE FROM N.A.				
RA	Yamada K., Chan W.-M., Chang C.-H., Dale J.M., Hsuan V.W., Lee J.M.,				
RA	Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,				
RA	Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,				

RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shim P.,  
 RA Southwick A., Tripp M.G., Wu T., Shinzaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.,  
 RT "Arabidopsis Full Length cDNA Clones",  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,  
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,  
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shim P.,  
 RA Southwick A., Tripp M.G., Wu T., Shinzaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.,  
 RT "Arabidopsis Open Reading Frame (ORF) Clones",  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB025633; BAA97244.1; -  
 DR EMBL; AF239719; AAF73960.1; -  
 DR EMBL; BT002944; AA022757.1; -  
 DR EMBL; BT004380; AA042374.1; -  
 DR InterPro; IPR005380; XS.  
 DR InterPro; IPR005381; zf.  
 DR Pfam; PF03468; XS; 1.  
 DR Pfam; PF03470; zf-XS; 1.  
 DR Hypothetical protein;  
 KW SEQUENCE 625 AA; 71971 MW; 456E2A1396706A96 CRC64;  
 Query Match 100.0%; Score 3313; DB 10; Length 625;  
 Best Local Similarity 100.0%; Pred. No. 1,7e-189; Indels 0; Gaps 0;  
 Matches 625; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSSRAGPMKSKKNVGGYRPEVOVLGAGTSLASSODDGEWEVTSKKNKPKNTSG 60  
 DB 1 MSSRAGPMKSKKNVGGYRPEVOVLGAGTSLASSODDGEWEVTSKKNKPKNTSG 60  
 QY 61 KTWVSONSNPPRAWGGOQGGHGSNVSGGNVSGKNGNGGICQANISGGRALSRKYDN 120  
 DB 61 KTWVSONSNPPRAWGGOQGGHGSNVSGGNVSGKNGNGGICQANISGGRALSRKYDN 120  
 QY 121 NVVAPPPVSRPPLREGMMWQANGSAOHTAVOEPDVEDVDNASEENDSDALDDSD 180  
 DB 121 NVVAPPPVSRPPLREGMMWQANGSAOHTAVOEPDVEDVDNASEENDSDALDDSD 180  
 QY 181 LASDDVDSVQSHSGSRKONKMKFKFGSLDLSISIEQINEPORQWHPACQNGGALDW 240  
 DB 181 LASDDVDSVQSHSGSRKONKMKFKFGSLDLSISIEQINEPORQWHPACQNGGALDW 240  
 QY 241 YNHLPLLAHARTGARRVYKLHRELAIVLEKDLQMGASVTPCGEITYGQMGGLGDEKDY 300  
 DB 241 YNHLPLLAHARTGARRVYKLHRELAIVLEKDLQMGASVTPCGEITYGQMGGLGDEKDY 300  
 QY 301 IYWPBWVITINWTRLDKDDNDKMLGMGNOELLEFDYKTEALRARSHSYGPGHSGSVLMFE 360  
 DB 301 IYWPBWVITINWTRLDKDDNDKMLGMGNOELLEFDYKTEALRARSHSYGPGHSGSVLMFE 360  
 QY 361 SSATGYLAERLHRELAEWGLDRIANGQKSMSSGGVROLYGFLATKODLDIFNHSOCK 420  
 DB 361 SSATGYLAERLHRELAEWGLDRIANGQKSMSSGGVROLYGFLATKODLDIFNHSOCK 420  
 QY 421 TRLKFELKSYOEWVYKELROISENOQNLNFKKLSKONKHAIVLESLEIMSEKLRRTA 480  
 DB 421 TRLKFELKSYOEWVYKELROISENOQNLNFKKLSKONKHAIVLESLEIMSEKLRRTA 480  
 QY 481 EDNRIVORTMQLHONREEMDAHDFPMDISIKOIHERRPAKEENPEMLQQORAVVGO 540  
 DB 481 EDNRIVORTMQLHONREEMDAHDFPMDISIKOIHERRPAKEENPEMLQQORAVVGO 540  
 QY 541 QOONINPSNDCKRAEVSFLEFQEKEMEEVEEREMLIKDOEKMEDMKKRHEET 600  
 DB 541 QOONINPSNDCKRAEVSFLEFQEKEMEEVEEREMLIKDOEKMEDMKKRHEET 600  
 QY 601 FDLEKPEDEALEQIMYKHLNEDD 625

DB 601 FDLEKPEDEALEQIMYKHLNEDD 625  
 RESULT 2  
 ID 07XY17 PRELIMINARY; PRT; 304 AA.  
 AC 07XY17  
 DT 01-OCT-2003 (T-EMBLrel. 25, Created)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Eml1.  
 OS Triticum aestivum (Wheat).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 CC Triticaceae; Triticum.  
 OX NCBI\_Taxid=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. PH 82-2-2;  
 RA Zhao X., Li Q., Zhang X.;  
 RT "Isolation and expression of a new kind of gene involve in  
 RT embryogenesis in Triticum aestivum L.",  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF542974; AAP80862.1; -  
 DR SEQUENCE 304 AA; 35958 MW; 89980215A9584208 CRC64;  
 Query Match 23.5%; Score 778; DB 10; Length 304;  
 Best Local Similarity 51.2%; Pred. No. 9,7e-39;  
 Matches 151; Conservative 82; Mismatches 82; Indels 4; Gaps 2;  
 QY 325 MGNQELLEFPDKYALRARSHSYGPGHSGSVLMFESSATGYLAERLHRELAEWGLDRI 384  
 DB 1 MGNQELLEFPDYATATARRHAYGPGHSGSVLMFESSATGYLAERLHRELAEWGLDRI 384  
 QY 385 AMGQKRSNF -SGVROLYGFLATKODLDIFNHSOCKTRLKFELKSYOEWVYKELROI 443  
 DB 61 TWONRRVPLPREGKQLYGFLARKEDMETFNRHOGKSLKYEMRSHENWVVAQMKOMSE 120  
 QY 444 DNQOLNFKKLSKONKHAIVLESLEIMSEKLRTRADNRIVORTMQLHONREEMDA 503  
 DB 121 DNQOLNFKKLSKONKHAIVLESLEIMSEKLRTRADNRIVORTMQLHONREEMDA 503  
 QY 504 HDRFPMDISIKOIHERRPAKEENPEMLQQORAVVGOQOONINPSNDCKRAEVSF 563  
 DB 181 OEFKFDQJENHKTEDKESFEFLDERAKA---RQCVDSGTTENRLRKEVQRF 237  
 QY 564 IEFQEKEMEEVEEREMLIKDOEKMEDMKKRHEETFDLEKPEDEALEQIMYK 618  
 DB 238 IECVADVQEFERHDEMIKAHEKKVQKKRYAKEVELEKREFDALTLGLMERH 292  
 RESULT 3  
 ID 07XY6 PRELIMINARY; PRT; 170 AA.  
 AC 07XY6  
 DT 01-OCT-2003 (T-EMBLrel. 25, Created)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE HORR (Fragment).  
 OS Triticum aestivum (Wheat).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 CC Triticaceae; Triticum.  
 OX NCBI\_Taxid=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. PH 82-2-2;  
 RA Li J.R., Wang F., Li Q.Z., Zhang X.S.;  
 RT "Gene isolation and expression of a new Zn-finger",  
 RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF469493; AAP80610.1; -  
 DR NON\_TER 1

SQ SEQUENCE 170 AA; 19980 MW; 0AFD21F1778B3068 CRC64;  
Query Match 16.6%; Score 548.5; DB 10; Length 170;  
Best Local Similarity 59.8%; Pred. No. 2.2e-25;  
Matches 101; Conservative 31; Mismatches 36; Indels 1; Gaps 1;  
QY 322 WLMGNOELLEYPDKTEALRARRSYGPOGRGMSVLMFESSATGYLAEERLHRELAEMGL 381  
DB 1 WLMGNOELLEYPDSYDAATKARAYAGPGCHGRGMSVLMFESSAVGWAEALHNFIDORT 60  
QY 382 DRANQKSMF-SCGYRQLYGLATKODLDINQHSQGTRLKFKELKSYQENVYKELRQ 440  
DB DRPTWQNRVRPPLPGGRQLYGLARKEDEMTNRRHCOGRSLKYEKRSHENEVVAQMKQ 120  
QY 441 ISEDNQOLNLYFKKLSKONKHAIVLESEIEMSEKLRRTAEDNRIRQR 489  
DB 121 MSEDNQOLNLYKKNKVKTEORSKVEETLITQKLRTEWENIYVASK 169  
RESULT 4  
ID 08VZ79 PRELIMINARY; PRT; 647 AA.  
AC 08VZ79;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN AT3G48670.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh U., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamlay A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh U., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamlay A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY065184; AAL38360.1; -;  
DR EMBL: BT000136; AAN15455.1; -;  
DR EMBL: IPR005379; XH.  
DR InterPro: IPR005380; XS.  
DR InterPro: IPR005381; zF.  
DR Pfam: PF03469; XH; 1.  
DR Pfam: PF03468; XS; 1.  
DR Pfam: PF03470; zF-XS; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 647 AA; 74824 MW; 4AE83BE6376279D0 CRC64;  
Query Match 7.3%; Score 242; DB 10; Length 647;  
Best Local Similarity 23.1%; Pred. No. 2.3e-06;  
Matches 111; Conservative 89; Mismatches 193; Indels 88; Gaps 19;  
QY 181 LASDDYDSVQSKSHSRKQKMKFKFPGSLDLSLEIQINEPQRMHCPACONGPAIDW 240  
DB 7 LSSDDSDSDISSES-----EMDEYGDKMVYLNKGGKTKVRLSPQ-AFICPYCPNKKKTSFQ 60  
QY 241 YNHLPLAARITG-----ARRVYKRELAELVEKDIQMKGASVIPCGETIYGQWKL 292  
DB 61 YK--DILQHASGVGNSNDKRSKAKERASHLALVKYLQODLADASAEAPSSKQKGNPF 118

QY 293 GEDEKDYIWPMPVILINTRLDKDNDKWLGMGNOL-EPYDK-YEALRARRSYGPOG 350  
DB 119 QODDHEKLVYPPKGIIVNIPPTKADGKSAGSSSKLDEYILRKFPNPRVPLNNYLG 178  
QY 351 HRGMSVLMFESSATGYLAEERLHRELAEMGLDIRAWGQRSMFSGVRLYGLATKODL 410  
DB 179 HSGTALVEFNDKNGNGLNGLFPDKAYTVDHGKQDKLK-----DGRKLGIYGIARADY 234  
QY 411 ---DINQHSQGTRLKFK-----ELKSYQENVYKELRQISEDNQOLNLYFKKLSKONKH 461  
DB 235 NGNNIIGENIRKTDGDKTALTEEARQOELLVONLRQVVE-----KKKD 281  
QY 462 AKVLESLEIEMSEKLRRTAEDNRIVRQRTQMGHQRSEMDADRPFMSIKOIHERRDA 521  
DB 282 MEKIEELCVKSEEL-----NQLMEKRNQOKHVR-ELMAIQETMSHIOCTIVD----- 330  
QY 522 KEENFEMLOOQERAKV-----GOOQONINPSSNDCKRAEVSFTLEPOE 568  
DB 331 DHEKLRLESERKLEIKCNELAKREVNIGTERMKL-----SEDLQNSKNS-LELA 385  
QY 569 KEMEVEVEREMLIKQEKMKMKRHHEIFDEKEDE-----ALEQL-----MYK 617  
DB 386 MEQKADEEVKKLAEDQROKEEL-----HEKILRLRORDQKQALEVEYQLGQNLVNRK 441  
QY 618 H 618  
DB 442 H 442  
RESULT 5  
ID 09SMN2 PRELIMINARY; PRT; 644 AA.  
AC 09SMN2;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN Tab19.180.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Choinsene N., Robert C., Brottier P., Wincker P., Catolico L.,  
RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X., Quelier F., Salanoubat M.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Arabidopsis sequencing project;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL133315; CAB62356.1; -;  
DR PIR: T46211; T46211.  
DR InterPro: IPR005379; XH.  
DR InterPro: IPR005380; XS.  
DR InterPro: IPR005381; zF.  
DR Pfam: PF03469; XH; 1.  
DR Pfam: PF03468; XS; 1.  
DR Pfam: PF03470; zF-XS; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 644 AA; 74411 MW; AFSNADA6F9811EB7 CRC64;  
Query Match 7.0%; Score 233.5; DB 10; Length 644;  
Best Local Similarity 22.9%; Pred. No. 7.4e-06;  
Matches 110; Conservative 88; Mismatches 192; Indels 91; Gaps 19;  
QY 181 LASDDYDSVQSKSHSRKQKMKFKFPGSLDLSLEIQINEPQRMHCPACONGPAIDW 240  
DB 7 LSSDDSDSDISSES-----EMDEYGDKMVYLNKGGKTKVRLSPQ-AFICPYCPNKKKTSFQ 60

```

QY 241 YNLPLLAHARTKG-----ARRVKLHRELAEVLKDLQMRGASVPCGEIYGQWGL 292
Db 61 YK--DLLQHAAGVGNNSNDKRSKAKESHALVKTLODPLADSASAEFSSKQKNGFI 118
QY 293 GEDDEKYEIVPMPVIMNTRLDKDNDKMLGNGOELL-EYFDK-YEALRAHSGYPOG 350
Db 119 QOCDBEKLVPMKGIIVNIPPTKADGRSAGSGSKLDEYILRGFNPTRVPLNNYLG 178
QY 351 HNGMVLMPSSATYGLAEERLHRELAEMGLDIRIAMGQKRSMSFGVRLYGLATKOL 410
Db 179 HSGTAIVERNKQWNGJHNLGLPDKATTVDGHGKMDLKK---DGPKLGLYGIATADY 234
QY 411 ----DINQHSOGKRLKF-----ELKSVOEMVYKELROISEDNOQLNFKNLSKONH 461
Db 235 NGNNIIGENDIRKTGDKLTIAELTEESARKOELLVQNLQVLEP-----KKD 281
QY 462 AKYLESELEMSKLRRTAEDNRIVRQRTQMOHEONREMDAHRFMDISIKOIHERRDA 521
Db 282 MKEIELGYSKSEL-----NQLMEKEKNQOKHYR-ELNATQERTMGIQKIVD---- 330
QY 522 KEENFMLOQOERAKTV-----GQOQONINSSNDCKRAAEVVSFIEFQE 568
Db 331 DHEKLRLESEKRLKIKCNELAKREVNQTERMTL---SEDLQONSKNS-LELA 385
QY 569 KEMEEFVEEREMLIKOEKMKEMDKRHHHEIFDLKEKPE-----ALBOL-----MYK 617
Db 386 MEQOKADEEVKCLADOREEL-----HEKTRLERQORDQKALIEVEQLKGLNVK 438
QY 618 H 618
Db 439 H 439

RESULT 6
Q9SBW2 PRELIMINARY; PRT; 629 AA.
ID AC Q9SBW2;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, last sequence update)
DT 01-OCT-2002 (TREMREL. 22, last annotation update)
DE Putative transcription factor X1.
GN P0663E10.7.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Lemont;
RA Chen M., Lucas J.R., Bennett J.L.;
RT "Different rates of divergence in SH2/Al-homologous regions of rice.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P0663E10.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
EMBL: AF101045; AAF21887.1; -.
EMBL: AP004317; BAB90725.1; -.
DR Gramene; Q9SBW2;
DR InterPro; IPR005379; XH.
DR InterPro; IPR005380; XS.
DR InterPro; IPR005381; zf.
DR Pfam; PF03469; XH; 1.
DR Pfam; PF03468; XS; 1.
DR Pfam; PF03470; zf-XS; 1.
SQ SEQUENCE 629 AA; 72715 MW; 4BB7928A7E3BFAF1 CRC64;

```

```

Matches 110; Conservative 96; Mismatches 179; Indels 137; Gaps 23;
QY 183 SDDVSDVSQSHSGRKQKMKFKFSGSLDLSIQINEPQOMHCPACQNPALDWN 242
Db 4 SDEDESELSDS-----EIDYADKPYARL-VAGEFKYDGSGYSCPL-SKKKKD-FN 54
QY 243 IHPLLAHTKGA---RRVK--LHRELAEVLKDL-----QMRGASVPCGEIYGQW 290
Db 55 INNLIQHASGVGAASNRRQAKDKATRALAKHKNLGLTSSGQSGQTAVEP----- 105
QY 291 GEDDEKYEIVPMPVIMNTRLDKDNDKMLGNGOELL-EYFDK-YEALRAHSGYPOG 350
Db 106 -QPLPRDEKFEVPMWGLVNVPTWKDG-RQIGSGHNLKEQLSFCLKTIPLMNFPG 163
QY 351 HNGMVLMPSSATYGLAEERLHRELAEMGLDIRIAMGQKRSMSFGVRLYGLATKOD- 409
Db 164 HSGNAIVERNKQWNGJHNLGLPDKATTVDGHGKMDLKK---DGPKLGLYGIATADY 219
QY 410 -----LDIFNQHSG----- 418
Db 220 TSPGLIGHLRKNGDKLTINDENGAARKTDLVANLANQIEVKRHLQELEVTYNERTT 279
QY 419 -----GK-----TRIKELKSVOEMVYKELROISEDNOQL-NYFRKLSKONKAKYLE 466
Db 280 SLEKMGQREQLQKYNEIRKMQQLAQHSGKIDENQKLSLELSKMSLEINTRSKEID 339
QY 467 ESLRIMSEKLR-----TAEDNRIVRQRTQMOHEONREMDAHRF 507
Db 340 E-IAKSDYDRRIIDOEKQKNAIKSHLKLATLEQRADENVLKLVERKKEKA- 394
QY 508 FMDISIKOIHERRDAKEMFEMLOOERAKVVGQOQONINSSNDCKR-RAEVSFIEF 566
Db 395 -YKLIKLEQOYDAKQK-LELDIQLKGLYVMKHM---PQEDSALKKIKIDELB-EL 447
QY 567 QEK--EMEEFVEEREMLIKOEKMKEMDKRHHHEIFDL 603
Db 448 QEKMDLDMESINQTLVYKERKSNTEMQDARKELENGLIDL 489

RESULT 7
Q9SA11 PRELIMINARY; PRT; 634 AA.
ID AC Q9SA11;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, last sequence update)
DT 01-JUN-2003 (TREMREL. 24, last annotation update)
DE F23A5.14 protein.
GN F23A5.14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S., Lee J.M., Li J.,
RA Gonzalez A., Liu A., Liu K., Sakano H., Koo T., Pham P., Vayenberg M.,
RA Huang B., Chin C., Choi E., Chlou J., Alatafi H., Brooks S., Chao Q.,
RA Conn L., Conway A., Hansen N., Johnson-Hopson C., Khan S., Kim C.,
RA Lam B., Nguyen M., Palm C., Shinn P., Tambunga G., Davis R.W.,
RA Ecker J.R., Federici N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F23A5 sequence.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Theologis A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
EMBL: AC011713; AAF14667.1; -.
DR PIR; E96840; E96840.
DR InterPro; IPR005379; XH.
DR InterPro; IPR005380; XS.

```

Query Match 6.54; Score 215.5; DB 10; Length 629;  
 Best Local Similarity 21.1%; Pred. No. 8.5e-05;

[illegible]

Query Match 6.3%; Score 208; DB 12; Length 1003;  
Best Local Similarity 18.4%; Pred. No. 0.00043;  
Matches 91; Conservative 104; Mismatches 199; Indels 100; Gaps 14

QY EFPPVEDVDVNASSEENDSDALDSDDDLASDDYSDVSQKSHGSKONKWFKFFGSID 212  
DB EEDDEEEDDEEEDDEEEDDEEDDDDED--NEDEDEDKKEDEBEDDGDGNK-----TIS 424  
QY SLSTLEQINPEPRONMHCPCQNGPCGALIMYNLHPLLAHARTGARGVKLHRE----- 263  
DB IQSSQSQQQQEPQQQ---EPQQQEPQQQEPQQQEPQQQEPQQQEPQQQEPQQQEPQQQ 481  
QY ---LAEVLEKDLQKRGASVIRCGEITYGQWKGLGEDEKDYETIWPMTIIMTRLDKDN 320  
DB EPQQQEPQQQDEQQQDEQ-----QQDEQQQDEQQQD-----EQEQQDEQQQD 523  
QY 321 KWLGMGNQELLEFDKTEALRAHRSYCGPQGRHGSVIMFESSATGYLAEARLHRELAE 380  
DB 524 E-----QQQDEQQQDEQDEQDEQDEQDEQDEQ-----EQEQEQELEQEQELEQ- 568  
QY 381 LDRIMAGCKSMFSGGVRQLYGLATKQDDLIIPNHSQKTRKLKELSYQEMVVKELRQ 440  
DB 569 -----EGELE---EQELEEQELEEQELEEQELEEQE 599  
QY 441 ISEDNQOLNYPKNTLSKONKAKVLESELSELTMSSEKLRRTAEDNRTVPRRTKQHQONBE 500  
DB 600 LEEBQELEBQELEBQELEBQELEBQELEBQELEBQEL---EBQEBELEBQEBELEBQEB 655  
QY 501 MDADHREFMDSIKOIHERRDAKEENFEMLQQOERAKVVGQOQONINPSSNDCRRABEV 560  
DB 656 LEEBQELEBQELEBQELEBQELEBQELEBQELEBQELEBQELEBQELEBQELEBQEB 715  
QY 561 SSFLEPQKHEEVEE-REMLINDQEKMDMKRHHBELFDLEKEDALEQ----- 614  
DB 716 EE-VEEBQEBQEBQEBQEBQEBQEBQEBQEBQEBQEBQEBQEBQEBQEBQEBQETV 774  
QY 615 ---MYKHGLHNED 625  
DB 775 EEPILIHGSSSEDE 788

RESULT 9  
Q8LSJ9 PRELIMINARY; PRT; 628 AA.  
AC Q8LSJ9  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
DE Line LH82 X1 (x1).  
GN X1.  
OS Zea mays (Maize).  
OC Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;  
OC Spermatophyta; Magnoliophyta, Liliopsida, Poales; Poaceae;  
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=line LH82;  
RC MEDLINE=21980573; PubMed=11959909;  
RX Yao H., Zhou Q., Li J., Smith H., Yandean M., Nikolau B.J.,  
RA Schnable P.S.;  
RT "From the Cover: Molecular characterization of meiotic recombination  
across the 140-kb multigenic al-sh2 interval of maize.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:6157-6162(2002).  
DR EMBL, AF44193; AAM2636.1; -.  
DR InterPro; IPR005379; XH.  
DR InterPro; IPR005380; XS.  
DR InterPro; IPR005381; zF.  
DR Pfam; PF03469; XH; 1.  
DR Pfam; PF03468; XF; 1.  
DR Pfam; PF03470; zF-XS; 1.  
QO SEQUENCE 628 AA; 71953 MW; 7B927B0FE03A37A CRC64;



DB 270 -----YTKVPENSKAYQSLPDLVQASREDLIIMPTVIIHNTATGKRKKGRAGELGNK 323

QY 329 ELLEYFDK-----YEALRRARHSYGPQGRHGMVLPESGATGYLAEARLHRELAEWGLD 382

DB 324 EM---DKTISELGFAGKSKSLYKRGHGLTLIFANSPLAGKEKERLADLIERODHG 379

QY 383 RIAMGQRSMFSGGV-----RQLYGFLATKODLDI FNOHSGKTRLK 424

DB 380 RIGWHARANQSGVNSPLVETDNRGTGKRRLIVGYLAISDMELDSKRRASLX 438

## RESULT 12

Q9S9P3 PRELIMINARY; PRT; 634 AA.

AC Q9S9P3; 01-MAY-2000 (TEMBLrel. 13, Created)

DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)

DT 01-JUN-2003 (TEMBLrel. 23, Last annotation update)

DE T24D18.1 protein.

GN T24D18.1.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC euroside II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxId=3702;

QX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Liu S.X., Yu G., Sakano H., Jhaveri A., Lee J.M., Lanz C., Pham P.,

RA Toriumi M., Chin C., Chlou J., Choi E., Chung M., Gonzalez A.,

RA Hong B., Koo T., Li J., Liu A., Vayberg M., Altafi H., Brooks S.,

RA Buehler E., Chao Q., Conn L., Conway A., Hansen N., Johnson-Hopson C.,

RA Khan S., Kim C., Lam B., Nguyen M., Palm C., Shinn P., Tambunga G.,

RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;

RT "The sequence of BAC T24D18 from Arabidopsis thaliana chromosome 1.";

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Theologis A.;

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC010924; AAF18488.1; -.

DR PIR; E86293; E86293.

DR InterPro; IPR005379; XH.

DR InterPro; IPR005380; XS.

DR InterPro; IPR005381; ZF.

DR Pfam; PF03469; XH; 1.

DR Pfam; PF03468; XS; 1.

DR Pfam; PF03470; ZF-XS; 1.

SQ SEQUENCE 634 AA; 72635 MW; 0071930DED7F41B0 CRC64;

Query Match 5.8%; Score 192; DB 10; Length 634;

Best Local Similarity 22.1%; Pred. No. 0.0022;

Matches 98; Conservative 80; Mismatches 179; Indels 86; Gaps 20;

QY 225 ONHCPCQNGRPAIDWYNLHPLAHA-----RTGGRVVKLHRELAEVLEKDLQMRGA 277

DB 38 QLRCPCCA-GKKKQD-YKTKELYAHAATGVSKSATSALOKANHLALAMLEL-AGTA 94

QY 278 SVLPCEIYGWKGLEDEKDYEIWPPVVIINNTRLDKDNDKMLGMGNQELLEVEFDKY 337

DB 95 ELPVPRPVVPRQDTEPRPNHNYVWPWVGIVNPL-KEADKELLDSAYVLTLSK 152

QY 338 EALRAHNSGPGQGRHGMVLPESGATGYLAEARLHRELAEWGLDRIAMGQRSMFSG-G 336

DB 153 KPLEVNAFWVEDQSIIVGLAKFGDWSGFAGATELKEFETQSSKKEWTER----SGDS 208

QY 397 VROLYGFATKODL-----DIFNOHSGKTRLKTELKSYOE--WVVEKLROI-----SE 443

DB 209 ESKAYWGCARADPFESQPIGELTSGEQLRTSDISQKNVQDRNTVLELSIMIAMTNE 268

QY 444 DNGQNLNFKNKS-----KQNGKAVKLESLSEIMSEKLRRTAEDNRIVQRKTMOH 494

DB 269 DLNKQVSYNRTAMSLQRLVLEKKNLHOAFADETKKQOMSLNHI---OKILYKEKLSN 325

QY 495 EONREEMDADHPFMDISQIHERRDAPKEENFEMLOOQERAKVVGQOONINSSNDPR 554

DB 326 ELDRKRRDESR-----AKOL-----EKHEALTELDROKI-----DEBK 359

QY 555 KRAEVSPIFEQEKMEBFVEEREMLIKOEKMEKRRHHEIFDLEKEFD--EALF 612

DB 360 RKSADANKSLQLASRQKRADESVLKLVBEHQROKEDALNK-----ILLERKQDLYQTLE 415

QY 613 -----QLMYKRGHINEDD 625

DB 416 MEIOELKGLQVW-KH-LGDDDD 436

## RESULT 13

Q815Y2 PRELIMINARY; PRT; 2760 AA.

AC Q815Y2; 01-MAR-2003 (TEMBLrel. 23, Created)

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Hypothetical protein.

GN PFL0315C.

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxId=36329;

QX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=3D7;

RA MEDLINE=22255705; PubMed=12368864;

RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,

RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,

RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,

Chan M.-S., Nene V., Shallom S.-J., Sun B., Peterson J., Angiuoli S.,

Raetea M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,

RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,

RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,

RA Venter J.C., Carnucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

RA Fraser C.M., Barrell B.;

RT "Genome sequence of the human malaria parasite Plasmodium falciparum.";

RL Nature 419:498-511 (2002).

DR EMBL; AE014845; AAN36152.1; -.

KW Hypothetical protein.

SQ SEQUENCE 2760 AA; 330896 MW; EE964C4845181AC CRC64;

Query Match 5.8%; Score 192; DB 5; Length 2760;

Best Local Similarity 19.7%; Pred. No. 0.014;

Matches 136; Conservative 142; Mismatches 272; Indels 142; Gaps 32;

QY 10 KEKNVGGYRPEYEOVLQGLAGTRLASSODDGEWEVISKQ--KNKPGNTSGKTVNSQN 67

DB 1473 KENDVNSKNERIEIVQVNEIKYTN--NKEEBG-----KKKILKEKEINDCLNDYINQ 1524

QY 68 SNPPRAMGGQOQGRGNSVGRGNVSGRGIGIOANISGRGRLSKRYONNFPAPPP 127

DB 1525 K-----KKEKKNNMAMTGRPIYKRONRNINIKNDLK--KYYSKSESGF----- 1568

QY 128 VSRPLEGGMNQARGSQAHTAVOEFPPVEDVDN--ASEEENDSDALDSD-----DDL 181

DB 1569 -----NDYAFALRFFPVITGVNSEPDYLDIDNQANSEKNDIHNHNNIKISKQK 1621

QY 182 ASDDYDSD-----VSQKSHGSKQNKWPFKFGSLDSLSEIQ 219

DB 1622 KENIYENSPHTYGRPIYKSKKNPNYNNKIKSTHNAIKKKRKTTLNKSISINSFTQK 1681

QY 220 NEPRQO-WHCPCQNGRPAIDWYNLHPL--LAA-----ARTGGRVVKLHREL----- 264

DB 1682 NSNANKIVKRTSICA--NTIDYNNSTYIKTIHKQONVEDQGYIDLTKRKLIYDLDEI 1739

QY 265 -AEVLEKDLQMGASVTPCGEIIYQ-----WKGLEDEKDYEIWPPVVIINNTRLDKD 318







**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:32:15 ; Search time 23 Seconds  
(without alignments)  
1402.880 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 3313  
Sequence: 1 MESSRAGPMSEKENVQGYRP.....EPDEALQQLMYKHGLNEDD 625

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: Issued Patents\_AA:\*  
2: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6A\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/6B\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/6C\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	182	5.5	1972	4	US-09-418-710-21 Sequence 21, App1
2	179.5	5.4	1388	2	US-08-685-576-1 Sequence 1, App1
3	173.5	5.2	1369	4	US-09-418-710-72 Sequence 72, App1
4	167.5	5.1	1898	1	US-08-056-200-94 Sequence 94, App1
5	167.5	5.1	1898	2	US-08-800-644-94 Sequence 94, App1
6	166	5.0	1162	2	US-08-728-323A-2 Sequence 2, App1
7	166	5.0	1162	4	US-09-298-568-2 Sequence 2, App1
8	166	5.0	1162	4	US-09-410-399-2 Sequence 2, App1
9	164.5	5.0	2662	4	US-09-585-684B-31 Sequence 2, App1
10	164	5.0	1354	3	US-08-685-871A-2 Sequence 2, App1
11	163	4.9	1388	2	US-08-685-576-4 Sequence 4, App1
12	162	4.9	966	4	US-09-688-188B-154 Sequence 154, App1
13	162	4.9	966	4	US-09-291-417D-154 Sequence 154, App1
14	161.5	4.9	1878	4	US-09-914-259-11 Sequence 11, App1
15	159.5	4.8	1886	4	US-08-938-105-3 Sequence 3, App1
16	159	4.8	816	2	US-08-533-306A-6 Sequence 6, App1
17	159	4.8	816	2	US-08-742-923A-6 Sequence 6, App1
18	159	4.8	885	2	US-08-533-306A-4 Sequence 4, App1
19	159	4.8	885	2	US-08-742-923A-4 Sequence 4, App1
20	159	4.8	911	4	US-09-688-188B-92 Sequence 92, App1
21	159	4.8	911	4	US-09-291-417D-92 Sequence 92, App1
22	158	4.8	1388	4	US-09-976-594-296 Sequence 296, App1
23	157.5	4.8	609	4	US-09-107-532A-4632 Sequence 4632, App1
24	157.5	4.8	1972	4	US-08-875-435B-3 Sequence 3, App1
25	156.5	4.7	683	6	5210183-3 Patent No. 5210183
26	156.5	4.7	976	3	US-09-104-324B-4 Sequence 4, App1
27	156.5	4.7	1939	4	US-09-310-187A-1 Sequence 1, App1

28	155	4.7	1285	4	US-09-976-594-507 Sequence 507, App
29	154.5	4.7	1972	4	US-08-875-435B-4 Sequence 4, App1
30	152.5	4.6	973	4	US-09-392-714-24 Sequence 24, App1
31	152.5	4.6	2954	4	US-09-150-867-11 Sequence 1, App1
32	151.5	4.6	1581	4	US-09-866-108A-15754 Sequence 15754, A
33	151	4.6	1244	4	US-09-543-681A-6274 Sequence 6274, Ap
34	150.5	4.5	2285	4	US-09-308-375-2 Sequence 2, App1
35	149.5	4.5	2101	1	US-08-466-390-4 Sequence 4, App1
36	149.5	4.5	2101	1	US-08-470-950-4 Sequence 4, App1
37	149.5	4.5	2101	1	US-08-467-781-4 Sequence 4, App1
38	149.5	4.5	2101	1	US-08-195-487-4 Sequence 4, App1
39	149.5	4.5	2101	2	US-08-483-924-4 Sequence 4, App1
40	149.5	4.5	2101	3	US-09-452-294-1 Sequence 1, App1
41	149.5	4.5	2101	5	PCT-US93-06160-4 Sequence 4, App1
42	146	4.4	1719	4	US-09-439-313-378 Sequence 378, App
43	146	4.4	1719	4	US-09-352-616A-378 Sequence 378, App
44	146	4.4	1719	4	US-09-636-215-378 Sequence 378, App
45	146	4.4	1719	4	US-09-685-166A-378 Sequence 378, App

## ALIGNMENTS

```
RESULT 1
US-09-418-710-21
; Sequence 21, Application US/09418710
; Patent No. 6596482
;
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418, 710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-09-418-710-21
```

Query Match 5.5%; Score 182; DB 4; Length 1972;  
Best Local Similarity 20.8%; Pred. No. 1.8e-05;  
Matches 113; Conservative 83; Mismatches 190; Indels 158; Gaps 20;

```
140 QARGSAGQTAIVQEPDVDDVDNASEENDSDALDDSDDLAS-----183
184 --DDVSDVSQSHGRKQKWKPFQSL--DSISIEINPQOWHCPACONGGAID 239
394 QRRGTSDSDIPSSKSDSDSEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 453
454 DDDDDQDSDSDSDTEGKTSMTKNTTSSVKSPSSSLTGHSPRNLIHAKGAPALC 513
240 WYNLHPL-----AHARTGARVRKLIHRELAELVKD-----LQWR 275
514 SSSQSAFLGTSSSTLTSSPHSGTSKRVRVTERELRIYLEGWQRETRIRNFGGLQE 573
276 GASVIFPGGITYQW-----KGLGDEKD-----YEIVW---303
574 VAVYAFPCGKKLQYBEVIKYLSRNGIMDISRNFSPSAKIRVGDYEAKDGPQEMQCLL 633
304 -----PWWITINTRLKDNDKWLGMGNQELLEFDKY---EALRARHSYVPGGRGWS 355
634 KEEDVIFRIAMEGRGRGRPNP-----DRRAAEBSRRRRRRKRPVAVGNA 679
356 VIMFESSA--TSYLEARLHRELAEMGLDRIAMQGRSMFSGVGVQLYGLATKODLDIF 413
```

```

Db      680 EFLDNADATLRLKLOAQLARQAOLIKLRLKLOKQOARVAKKQQAIVMAEE----- 734
Qy      414 NQHSQKTKLKFELK-SYQEMVVKELROI SEDN-----QOLNYPKXLSKONKHAUVLEES 468
Db      735 -----KXKQKQIKIMKQOEKIKRIQOIRMEKELRAQOILKAKKKKEEAMANKLLEAE 788
Qy      469 LEIMSEKLRRTAEDNRIIVORTKMOHQREEMDAHDFMDSIKOIHERRDAKEENFEM 528
Db      789 KRIRKEMER-----QOAVLKH-QERERRROH-----MMLMAMERKKAEEK--ER 833
Qy      529 LOQOERAKVVGQOQONINPSSNDDCKRAEBVSVFLEQKEMEFEVEREMLIKDOEK 588
Db      834 LKQEK-----DEKLN-----KEXLEORLELEW-AKELKKP 866
Qy      589 MEDM 592
Db      867 NEDM 870

RESULT 2
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozo
; APPLICANT: Iwanatsu, Akihito
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-685-576-1
Query Match      5.4%; Score 179.5; DB 2; Length 1388;

```

```

Best Local Similarity 20.8%; Pred. No. 1,86-05;
Matches 116; Conservative 108; Mismatches 182; Indels 151; Gaps 27;

Qy      147 QHTAVQEPFDDVDVNDNASEENDSDALDDSDLDASDDVSDVSQSHSGRKONKMKFK 206
Db      509 QHKMAEYQKADHEADKKNLENDVNSIKDQLEDLK-----KRNQ----- 549
Qy      207 FFGSLDLSLEQINPQRMHCPACONGPAIDWNLPLLAHARTGAR----- 256
Db      550 -----SQISTEKVNQLORO-----LDEN--ALRTESDTRARLKTOAESK 590
Qy      257 -----RVKLHRE--LAEVLEKIQMR--GASVPCGEIYQW 289
Db      591 QIQOLESNNRDLDQKNCLETAKLKEEFINQSVLESERDRTHGSEII--NDLQRI 648
Qy      290 KGLGDEKDYELVPMVVIIN-----TRLDKNDNDKLGWG-----NOELLEYPDKY 337
Db      649 SGLBEDYKNGKILLAVVELEKQLOERFTDLEKKNNEIDMTYOLKVIQSLBGEETEH 708
Qy      338 EALRAHSHVSGPOGRGWSVLMFESSATGYLAEERL--HRELAEWGLDIRIANGQKR-SMF 393
Db      709 KATYRLADKNTYESIE-----EAKSEAMKEMEKKLESERTIKQKVENLLEAKRCSTL 764
Qy      394 SGGRQLYGFLATKQDI-----DIPNQSCKTKLKEELKSYQEMVVKELROI SEDNQ 446
Db      765 DCDLKQ-----SQQKINELLKQKDVINEDVRNLT-LKTBQETQKCLIQ--NDLKQOTQ 815
Qy      447 QLAYFK--NKLKONKHAUVLEESLEIMSEKLRRTAEDNRIIVORTKMOHQREEMDA 503
Db      816 QVNTLKSEKOLKQENHLLKMSLEKQNAELRKE-----ROADQMKELQPOLBA 868
Qy      504 HDRF--FMDSIKOIHERRDAKEENFEMLOOQ-----EAKVVGQOQONINSSND-C 553
Db      869 EGYFSTLYKTQVRILKECEKEKTKLCKELOQKQLOQEDRSDLAQULEITLTADSEQLA 928
Qy      554 KRAAEVSVFLEQKEMEFEVEREMLIKDOEK-----KMEWMKKRHHHEIFDL- 603
Db      929 RSIABEYSDLE-KEKTKKE-LEIKEMARHKQLTETKATIASLEETNTLTSDVANILA 986
Qy      604 -EKE-----FDEALEQL 614
Db      987 NEKEELNKKLKEAQEQL 1003

RESULT 3
US-09-418-710-72
; Sequence 72, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1669
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-418-710-72

Query Match      5.2%; Score 173.5; DB 4; Length 1669;
Best Local Similarity 20.1%; Pred. No. 8,6e-05;
Matches 128; Conservative 94; Mismatches 213; Indels 203; Gaps 25;

Qy      47 ISKNKKNKPGNTSGKTWVSQNSNPPRANGQOQGRGSNVSGRGNVNSGRGNNG----- 100

```

```

Db 341 LSTGRRFFGN---QTPWMSAPIL-----HQGKKAVSNMNVNPKYQTHHSHPSLIVE 393
Qy 101 --RGIQANISGRGALSRKYDNFVAPRPVSPLEGGMNWQARGGSAOHTVAQOEPDVE 158
Db 394 QFRGTSDI-----PSSK-----DSEDSNEDEEDDEE 421
Qy 159 DDVUNASSEENDSDALDSDS-----DDDLASDDYDSQVSKHSGSKKUNKMFK 205
Db 422 EDEEDEDSDSDSQSESDSNSDSESDTEGSESEDDDDKXODESDSD---TEGEXTSMKLNK 477
Qy 206 KFFGSLDLSISIQINIEPORQWCHPACQONPGAIIDWNLHPLL-----AHARTKG 254
Db 478 TTSSSPKMSLGHSTPRYLHIAKAPGAPALCSCSGPAPFGTSSSTLTSPPHSGTSK 557
Qy 255 ARVYLARELEVELEKD-----LQWRGASVIPCGEIVGOW-----KGL 292
Db 538 RRVYDERELRPLFVGMQRETRINFGRLQGEVAVYAPCGCKLQYEVIKYLSRNGI 597
Qy 293 GEDEKD-----YELVW-----PBYVILNTRLKDODNDKWLK 324
Db 598 MDISGDNFSFSAKIRVGDPEYARDDGPOEQWQWCLLKEEDVI PRIRAMEGRRGPPNP--- 653
Qy 325 MGNQELLEYFDKY---BALRABHSGYQGHGRGWSVLMFSSA---TQYLEARHARELAEM 379
Db 654 -----DQORAREBSKMRMRKRPRPVGNAEFLDNADALRLKLQOEA-RQAAQI 702
Qy 380 GLDIRIAMGQKMSFSGVRYQLYGLATKODLDI FNOHSGQKTRLKFELKSY--QEWVVKEL 438
Db 703 KLRRLQOQEOQARVAKKQAQIAMAEB-----KKQKQKQIKIHQOQEKIKRI 751
Qy 439 QOISEDN---OQNLVFNKLSKQNKHAKVLEBSLEIWSKLRRLAEDNRIVYQRTKMQH 494
Db 752 QOIRKEKELRAQOQIIEAKKKKKEBANATLIEAKRIKEREMR-----RQQAALLK 802
Qy 495 EONREMAHDFPMFDSIKQIHERDAKENEEMLOQOQERAKVYGOOQONINSSNDCCR 554
Db 803 ROERRRRROH---WMLMAMEARKKAERK-ERLKQER-----DEKRLN----- 842
Qy 555 KRAEVSSEFIEQEKEMEEFEVEREMLIKDOEKQMEDM 592
Db 843 -----KERKLEBORLELEM-AKELKKPNEDM 867

RESULT 4
US-08-056-200-94
: Sequence 94, Application US/08056200
: Patent No. 5616500
: GENERAL INFORMATION:
: APPLICANT: Steinert, Peter M.
: APPLICANT: Lee, Seung-Chul
: APPLICANT: Kim, In-Gyu
: APPLICANT: Chung, Soo-Il
: APPLICANT: Park, Sang-Chul
: TITLE OF INVENTION: Trichonyalin and Transglutaminase-3 and
: NUMBER OF SEQUENCES: 117
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 620 Newport Center Drive, Sixteenth Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: U.S.A.
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/056,200
: FILING DATE: 30-APR-1993
: CLASSIFICATION: 435

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Pedrick, Michael F.
: REGISTRATION NUMBER: 36,799
: REFERENCE/DOCKET NUMBER: NIH054.001A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (714) 760-0404
: TELEFAX: (714) 760-9502
: INFORMATION FOR SEQ ID NO: 94:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1898 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-056-200-94

Query Match 5.1%; Score 167.5; DB 1; Length 1898;
Best Local Similarity 20.9%; Pred. No. 0.00024;
Matches 98; Conservative 85; Mismatches 162; Indels 123; Gaps 22;

QY 246 LLAHATKAGARVYLHRELAEVLEKDIQ----MGASVLPCCGIIYGGWKGLGEDEKDYE 300
DB 1140 LLIREEREK-RRKQELRYQRE--EEELQQRKQRYDEDDRSRLAKQW----EPKXNA 1192
QY 301 IWPWPVILINMTRLDKDNDKMLGNGOELLEFYDKYEAUR-----ARHSYGPCGHRG- 353
DB 1193 V-----RDKVYCKGRNEQFRQLEDSDQVRDROSQDULQHLGSGQERDRE 1238
QY 354 MSVLMFSSATGYLAEERLHLEAEMGLDRIAMQ-----KSMFSGGVQ 399
DB 1239 QERRRWQDANHFPFEEQLEBEQEKAKRRDRKSQBEKQLLREEREKRRQGTDRKPRE 1298
QY 400 LYGLATLKQDLDIDNQ-----HSQKTRLKFE-----LKSQYQEMV 435
DB 1299 EEOQLQREBPQLRQERDRKFRFEEELHOGGRKFLFEEQRLNREERKFLKEEQQLRL 1358
QY 436 KELRQIED-----NQQLN-----YFNKLSKQNHAKVLEESLEINSEKLR 478
DB 1359 EEBQLQRODRKFRFEEQQLSRQERDRKFRFEEQOVVRQDERKFLFEEQQLQOERHRK 1418
QY 479 TAEINRIYRQRTKMQ-HEQNREEMDAHDFPMDSIKQI-HERDPAK-----EEN 525
DB 1419 FREFEQQLQREBEQQLRQER-----DKFLFEEQQLRQERDRKFRFEEQLRSQEPFRK 1472
QY 526 F---ENLQOQERAKVVGQOQINPSSNDCKRAEVSFIPEQEKEMEFVEEREML 581
DB 1473 FLFEEQQLHRQQRKFLQEBQQLRQERQ-GRQORDRKFR--EBQLQREBEQQLS 1529
QY 582 IKDOEK--MEDMKGRHHEIFLDEKFFDELQEDLMTKRG---LHNEP 624
DB 1530 RQERDRKFLFEEQKYRQEQ---ERRKMEDEQQLRQEGQOQLROED 1573

RESULT 5
US-08-800-644-94
: Sequence 94, Application US/08800644
: Patent No. 5958752
: GENERAL INFORMATION:
: APPLICANT: Steinert, Peter M.
: APPLICANT: Lee, Seung-Chul
: APPLICANT: Kim, In-Gyu
: APPLICANT: Chung, Soo-Il
: APPLICANT: Park, Sang-Chul
: TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
: NUMBER OF SEQUENCES: 117
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 620 Newport Center Drive, Sixteenth Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: U.S.A.
: ZIP: 92660
: COMPUTER READABLE FORM:

```





```

; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRN
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-2

```

Query Match	5.0%;	Score	166;	DB	4;	Length	1162;
Best Local Similarity	19.6%;	Pred.	No.	0.00016;			
Matches	46;	Conservative	66;	Mismatches	85;	Indels	38;
						Gaps	5

[illegible]

```

RESULT 8
US-09-410-399-2
; Sequence 2, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
; US-09-410-399-2

```

	Query Match	5.0%	Score 166;	DB 4;	Length 1162;
	Best Local Similarity	19.6%;	Pred. No.	0.00016;	
	Matches	46;	Conservative	66;	Mismatches 85; Indels 38; Gaps 5
OY	407 KODIDIFNHSOGKTRFLKFELKSVOEWMVKLRQISEDNOQLNYFRKLISKONKHAKYLE	466			
Dd	735 ::::~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~:	794			
OY	467 ESLFIMSEKTRLTLDNRIVRQTGNQKNREEMDADRPFMSIKOINHERDAKENFE	526			
Dd	795 QELEBOBEL-----EEBOLEBEBOLEBEBOLEBEBOLEBEBOLEBEBOLE	836			
OY	527 EMLOQEERAKVVGGOQNINPNSNDCKRAEV---SSFLFOEKEMEVEFEREMI	582			
Dd	837 QELEEEO----VEEQEEVEEBQOEBOBLEVEBEOBEOBEOBEOBLEVEBQOEOL	892			
OY	583 KD-----QKKMKDKNNHRIFFDLKEFDALTEL-----MYGGLINNED	625			
Dd	893 EEVEBOEOLEVEBOEEOLEVEBOEEOGVBOEOETVEEPILIGSSED	947			

RESULT 9  
US-09-595-684B-31  
; Sequence 31, Application US/09595684B  
; Patent No. 6544766  
; GENERAL INFORMATION:

```

1  ? APPLICANT: Beraud, Christophe
2  ? APPLICANT: Ohashi, Cara
3  ? APPLICANT: Sakowicz, Roman
4  ? APPLICANT: Vaisberg, Eugeni
5  ? APPLICANT: Wood, Kenneth
6  ? APPLICANT: Yu, Ming
7  ? TITLE OF INVENTION: Human kinesins and methods of producing
8  ? TITLE OF INVENTION: and purifying human kinesins
9  ? FILE REFERENCE: cytopo036
10 ? CURRENT APPLICATION NUMBER: US/09/595,684B
11 ? CURRENT FILING DATE: 2002-06-24
12 ? PRIOR APPLICATION NUMBER: 09/235,612
13 ? PRIOR FILING DATE: 2000-04-20
14 ? NUMBER OF SEQ. ID NOS: 105
15 ? SOFTWARE: FastSeq for Windows Version 4.0
16 ? SEQ ID NO 31
17 ? LENGTH: 2662
18 ? TYPE: PRT
19 ? ORGANISM: Human
20 ? US-09-595-684B-31

```

Query Match	5.0%;	Score 164.5;	DB 4;	Length 2662;
Best Local Similarity	18.2%;	Pred. No. 0.0067;		
Matches 139;	Conservative 134;	Mismatches 271;	Indels 221;	Gaps 30;

```

OY 18 YRPEV-----EOLVGLAGLRTIASSODG-----GEMVYISKNNKNGKNGTSGKTWVSQNS 68
Dd 348 YRKIMOLKXQLEBVSLETRQAOMEKQOLQLEEKDOLKQVQNEKXENLIRMLVTSSSL 407
OY 69 NPPRWMGQOQGRGSNVSGRGNNVSGGNGNGRCIQANISGRGALG---RKYDNNFVA 124
Dd 408 TLQOLAKRKRRTVWICIGKINRKNNSNYADQFNIPINITTKTKLINLIREIDESVCS 467
OY 125 PEPVSRPEL-----EGGWNMOARGSAOHTAVQEPEDVEDVDN-----163
Dd 468 ESDVFSNTLDTLSIEIWM-PATKLLNDBNIESIANSIRADYDNLVDLYEQURTEKEMEL 526
OY 164 ASEBNSDSDALD-----DSDDDLA-----SDDYSDVSQKSHGSRK 199
Dd 527 KJAKKNDLDEFALERKTKKQDOEQMLIHEISNLKVLWGHREVVQDLENELISSKVELLRE 586
OY 200 QNKKFKPFPSGLD-----SLSEIQNEPQOROMHCPACQNGCALDWNVLHPLLA 248
Dd 587 KEOIKLQEQEYISQKLENIKMDISYSLESEIDEPKQKQ-----TLFAETVAL 635
OY 249 HARTKGA-----RKYCLHRELAEV-----LEKDQMGAGAVIPCGEYIGQWKGJGEDEKD 298
Dd 636 DAKESAFSLSENELKEMKELATTKYKQEMNDIQLVQS-----QLEAKKMQVDLEKE 689
OY 299 YEIYW-----PPIVITMTRLDKODNDKMLGMCNOGBLAEYFDKYLELAPRHS 345
Dd 690 LOSAFNEITKLTSLIDKVPDDLCLCNLEBEKTYDL-----QKYLNEVEENELARE-- 741
OY 346 YGPGHGRGMSVITPESSATGYLAEARLHRELAEWGLD-RIAMQGRKMFSGGV---OLY 401
Dd 742 -----EYILSELKSLPSEVERARKEIQDSESELHIITTSKDLFSPVWAKESVQ 792
OY 402 GFL---ATKODL-----DIFNQS 417
Dd 793 GLLEIEIKTKDIDLATTSQNYKSTDOEFQNFKTLMDPEQKYKVLLEENERNKMOEIVNLK 852
OY 418 QG-----KTRLKPELKSVOEVMVKELROIASENOOLNFPKGLSKONKFAKYLEE 467
Dd 853 EAQKFDSSLGAKTKELSYKQEOLEO---KTRVEQELNEMEQAKEOLENRDLSQIYER 908
OY 468 SLEIMSEKLRTIADNRIVRQ-----RTKQHEQNRREMDAHRPFMDISIKOIHERR 519
Dd 909 EKLITITEKLOQTLLEVYKTLIOEKDOLQOLBSQIOERDOLKSJDHIDHTVNM-NIDTOELR 967
OY 520 DAKENEMEMLOOE-----RAKVVGQOQOQNNIPESNDCKRABEVSSFTFOKEKMEEF 574
Dd 968 NA-----LESLKQETITWTLKSKEISSEVSRNLNHEENGTETKD-----EFQOK-MVGI 1014

```

QY 575 VEREWLIDKOEKMEKMDK-----RHHEIFDLKEFDEALQOLM 615  
DB 1015 DKKQDLKAKTQTLTADVDNNEILKQKRIKFILOEKNE-LQOML 1058  
RESULT 10  
US-08-685-871-2  
Sequence 2, Application US/08685871  
Patent No. 6013499  
GENERAL INFORMATION:  
APPLICANT: NARUMIYA, Shuh  
TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,871  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-184102  
FILING DATE: 25-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-262553  
FILING DATE: 14-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16887/845  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1354 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-685-871-2  
Query Match 5.0%; Score 164; DB 3; Length 1354;  
Best Local Similarity 21.3%; Pred. No. 0.00029;  
Matches 118; Conservative 89; Mismatches 201; Indels 146; Gaps 27;  
QY 147 QHTAVQEFPPVEDVDNASEENDSDALDDSDDDASDDVDVSQKSHGS----- 197  
DB 492 QHRINVEYORKAEQNEKRRNVENSVTLKQLEDLKYSQNSQLANEKHSQQLKQLEBAN 551  
QY 198 -----RKQNKWFKKFGSLDSISTEQINEPQFQWCHCPACQNGPAGIDPYNH 244  
DB 552 DLRTRESTAVRLKSHTEMSKISQLESIN-RELQERNRILLENKSQTDK---DYTQLO 607  
QY 245 PUL-AHARTK-----ARRVKLHRELAVLELDQMRGASVAPCEIYGQWKG-- 291  
DB 608 AILAEARRDRGHDSMIGDLQARTSLQEEVGH-LKHNLE-----KVGSRKEAQ 656  
QY 292 --LGDEVDYELVWPVYIMNTRLDKDNDKMLGNGOELLLEFYDYKALRAHSHYGPQ 349  
DB 657 DMLNHSKE-----KNLEIDLNKTKSL-QQRLQGVNHEKVKYKALTDKHQ 703  
QY 350 G-HRGMSVLMPESSATGYLAEARLHRELAENGDLRIANGQR-SMFGGVRQLYGLATK 407

DB 704 SIEAKSVAMCKMEK-LKEERAREKAE---NRVQIEKQCSMLDVLQKQ-----SQ 752  
QY 408 QDLIDFNHSG-KTRLKFKELKSYOEYVVKELRQISEDNQOLNFKKLSKONKHAYLE 466  
DB 753 QKL-----HILGNKRMEDEVKN-----LTLOLEQ--ESNKL-LLOVELKTOAFEDNLK 801  
QY 467 ESLEIWEKLTARTADNRIVR-----QRTKM-----QHEQNRREMDADRF--FMDSIK 513  
DB 802 GLEKQKQKQEIINTLLEKRLLEPQLTKQKRGNGQWRBELQDQLEAQQFSTLYKTOVK 861  
QY 514 QIHERDAR-EENFEMLOQ-----OERA 535  
DB 862 ELKEHELKKNRELLKQIOLQNEKETLATQDLAETKASSQLARGLEBOYFELTQSK 921  
QY 536 KVVGOQONINPESNDCKRAEVSFTIEQEKEMEVEVEREWLIDKOEKMEKMDK 595  
DB 922 KAASRRNOEL-TDKHTVSRLEANSML---TQIETLRNEELTQKMKAEVEYLE 976  
QY 596 HHEIFDLKEFDE 609  
DB 977 KEELISNLKAFK 990

RESULT 11  
US-08-685-576-4  
Sequence 4, Application US/08685576  
Patent No. 5906819  
GENERAL INFORMATION:  
APPLICANT: Kaibuchi, Kozi  
APPLICANT: Iwamatsu, Akihito  
APPLICANT: Nakano, Takeshi  
APPLICANT: Ito, Masaaki  
APPLICANT: Takahashi, No. 5906819uaki  
TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,576  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-325129  
FILING DATE: 20-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-17150  
FILING DATE: 05-JAN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-131206  
FILING DATE: 26-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16887/843  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1388 amino acids  
TYPE: amino acid



;; SEQ ID NO 154  
;; LENGTH: 966  
;; TYPE: PRT  
;; ORGANISM: Murine sp.  
US-09-291-417D-154

Query Match 4.9%; Score 162; DB 4; Length 966;  
Best Local Similarity 18.3%; Pred. No. 0.0026; Indels 256; Gaps 32;  
Matches 138; Conservative 121; Mismatches 238;

```

19 REVEEVLQVGLAGTRLASS-----QDDGEMWEVI-----47
Db 282 RPSAQLQLQPPVSRVTSKALRELYAEAKAEVMEIEEDRGDEBEDAVDAVPLVNH 341
Qy 48 -SKNKAKPQNTSGKTWVSNPNPRAWGQOQSGSNVSGKANNVSGNGNGGION 106
Db 342 QDSANVTQPLDSNK--LLQDSSSTPLPSQPE---PVNGPSQPSGQ---PLQTT 390
Qy 107 ISGRGRLSKYDNNFAPPPV--SRPLEGGMWQARAGSAQHTAVOEPPVEDVDNA 164
Db 391 SPADGLS---KNDNDLVVPLKSRP-----LSMDAR---IQMDEKQIPD-----431
Qy 165 SEENDSDALDDSDDLASDDYSDVSQKSGSRKQNKFFKFGSL---DSLSIQINE 221
Db 432 -QDENPSPA-----ASKSQANOSRPNSSALETLGEGALTINGLELSSVT 476
Qy 222 POROMHCPACON-----GPGALDWMYHLPLAHARTKARVYLHRELAVALKOL 272
Db 477 PSHSKRADCSNLSTESMDYGTSLADSLNKGTSLSLKGS---KLH-----NKTLL 526
Qy 273 QMRGASVIPCSEI--YGQWKGLEDEKDEYIWPMPVITINTRLDKDNNDKMLQMGNE 331
Db 527 KTRRFVVDGVEVSTTSKIISEDEK-----KDEEMFPL-----560
Qy 332 EYFDYKYLARHSGYFGQHGKMSVLMFESSATGYLEAERLH-----RELAEMGL 381
Db 561 ---RROELRELRLQKEHRENOTL---SKHELOLEQNHKRFEOEINAKKKFYDEL 612
Qy 382 DRIAMQKRSKSM-----FSGGVOLYGLATKODLIFNOSHOGKTRLKFELKSYQEMV 435
Db 613 ENIERQOQKQOEVKEMEDHVRKREKAKRIRLEODRDYAKFOEOLK--QMKKVKSEVEKL 671
Qy 436 KELRO-----ISDNQOLVYFNKLSKQNHAKVLEESLEIMSE 474
Db 672 RQOKESMKQKMEHSHQKORLDRDFAKQKEDLELAKLITTEKR--FIDCKERDCLSK 730
Qy 475 K---LR-----RTAEDNRIVRORTYQO-----HEONREMDAHRFF 508
Db 731 KQELNDREALWEMEHQLOERHQVLVQOLKQOYFLQHDLLRKHKEKREQEMQRYNORM 790
Qy 509 MDSI-----KQIH-----ERRDAKE 524
Db 791 MEQIKVQOQOKARLPKIQRSDETRMAMWKSLHNGASASQOEKTKQOSQOEKQ 850
Qy 525 NFEMLQOQOKAK-----VVGQOQON---INPSNDCKRAEVSFIPOKEWEEFYE 576
Db 851 KAEELQOQOKHEHQMRQWVAQCESNMSHELQOLNENKCYLLVEHETQKLALDESHQSIK 910
Qy 577 EREMILKQOQKQMED---MKRRHHEIFDLEKE 606
Db 911 EWRDKLRPKKALBEDLNQKREQEMFKLSE 943

```

RESULT 14  
US-09-914-259-11  
Sequence 11, Application US/09914259  
GENERAL INFORMATION:  
APPLICANT: Makowski, Lee  
APPLICANT: Hyman, Paul  
APPLICANT: Williams, Mark  
TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES  
FILE REFERENCE: 8471-010-999

;; CURRENT APPLICATION NUMBER: US/09/914,259  
;; CURRENT FILING DATE: 2000-11-21  
;; NUMBER OF SEQ ID NOS: 180  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 11  
;; LENGTH: 3878  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-914-259-11

Query Match 4.9%; Score 161.5; DB 4; Length 3878;  
Best Local Similarity 19.0%; Pred. No. 0.002; Indels 179; Gaps 26;  
Matches 111; Conservative 101; Mismatches 194;

```

163 NASEENDSDALDDSDDLASDDYSDV--SOKSHGRKQNKFFK-----207
Db 1668 NENQGEVEEQTFFKEKLDKPEVPELISNERYALQANNRLITLLEVYTKTAAVEET 1727
Qy 208 ---FCSLDLSIEQINEPOROMHCPACONGPAIDWYTNLHPLAHARTKARVYK--L 260
Db 1728 IGRHVLGIIDRSSKQ-----SSASLIW-----RSEKASVYKCV 1762
Qy 261 HRELAVALKOLQMRGASVIPCSEIYGQWKGLEDE--EKDYIWPMPVITINTRLDKOD 318
Db 1763 HEHTRVTDISIPSYSGSDMPRNDI--NMMSKYTEBETLSQRLVNSGFA---GTEIDPEN 1818
Qy 319 NDKWLMGKN-----QELTFYFDKYELARHSGYFGQHGKMSVLMFES---SATGYL 367
Db 1819 EELMNLISSRLQAAVEKLELAISETS--QLEHAKVTO---TELMRESFROKQKATBSL 1872
Qy 368 EA---ERLHRELAEMGLDRIAMQKRSKSMFSGGVOLYGLF--ATKODLIFNOSHOGKT 421
Db 1873 KQOELRERLHEE-----SRARQALAYELSGAEVIDGYADEKTLFROIQEKT 1921
Qy 422 RLKPELKSQYQEMVYKELROISEDNQOLYFNKLSKQNHAK-----VIEESLEIM 472
Db 1922 DIIDLEQELLCASNRLOELAEBOQIQEERELLSRQKEMKQKAGVEQOLLOETKLM 1981
Qy 473 SEKLRTAEDNR-----VRORTYQOHEONREMDAD-----505
Db 1982 KELEVOQCAEKVRRDLOKQVLALEIDVEQVSRFIELEQKTEMLMDLQONALEKOL 2041
Qy 506 ---RPFMD--SIKQIHERDAKENFEMILQOQERAKVYQ-----QQONINSSND-- 551
Db 2042 EKMRKFLDQALDREH--RDVPQOEIQLEQO--LKVPRPQPISEHQTRVEQJANHLK 2098
Qy 552 -----DCKRAEVSFIPOKEWEEF-----VEER-----578
Db 2099 EKTDKSEILLSKEQLORDIOERNEIEK--LEFRVRELOALLYSADTFQVYEDRKIFGA 2157
Qy 579 ---EMILKQOQKQMEDKKKHHEIFDLEKEPDEALQOLMYKH 618
Db 2158 VEAPELISLEVQLOAERDALDKKEKQITNLBEQLOQFPELEENKN 2202

```

RESULT 15  
US-08-938-105-3  
Sequence 3, Application US/08938105  
Patent No. 6353151  
GENERAL INFORMATION:  
APPLICANT: Weinand, Leslie A.  
APPLICANT: Viketiom, Karen L.  
TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: CO  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,105  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Crook, Mannell M.  
REGISTRATION NUMBER: 31,071  
REFERENCE/DOCKET NUMBER: 3595-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1886 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-938-105-3

Query Match 4.8%; Score 159.5; DB 4; Length 1886;  
Best Local Similarity 19.3%; Pred. No. 0.001;  
Matches 112; Conservative 106; Mismatches 193; Indels 169; Gaps 29;

QY 139 WQARGSAOHTAVOEPFVDVVDVNASEEND--SDALDGS--DDDLASDDYDSVGKSH 195  
DB 1248 WQTRKGLST--QOMEDLRQLBEGKAKNALAHQASARHDCDLIREQYEEMAKAE 1305  
QY 196 GSRKQK-----WFKKFFGSLDLS--IEQINEPOROWHCPACONGPAIDWYNLHPL 247  
DB 1306 LQRLVLSKANSEVAQWRTKY--ETDAIQRTLELEAKKL--AQRLQDAEBAEAVN----- 1357  
QY 248 AHARTGARVRVKLHRELAEV--LEKDLQMGASVIFCGEITYGQKIGEDKDIYIWP 305  
DB 1358 --AKCSLEKTK--HRLQNETIEDLVVERSNAAA-----AALDKKQNFDKI--- 1401  
QY 306 MVIMNTRLDKDNNDKMLGNGNOLLEYPDYKALRAHSYGPGRGMSVLMF----- 359  
DB 1402 -----LAEWKQYEESSQSELESQKEARSISTELFKLNAY 1437  
QY 360 ----ESSATGYLEAEHLHRELAEMGLDRIAMGQK-----RSMFGGVQLYGFATK 407  
DB 1438 EBSLEHLETFKRNKQLQBERISDL--TBQLGEGGNVHELEKTRQLEVEKLEISAL--- 1493  
QY 408 QDLIDFNQHSQK--TRLKTELKSYQEMVVELRQISEDNOQLNYFRNKLKSKQNHAKV-- 464  
DB 1494 BEAFASLHEHEGKILRAQLEFNQIKAEIERKLAEKDEMEQ-----AKRN--HLRVVD 1544  
QY 465 -LEESLEIMSEKIRTAEDNRIYRQRTKQHEQNREEM-----DAH 504  
DB 1545 SLQTSILD-----AETRSREALRVKKMGDLNEMEIQLSQANRIASEAQKLNQAHA 1598  
QY 505 DREFFM-----DSIKQ-----IHERRDA-----KEENPEMLQOQERAKVVGQO--- 542  
DB 1599 LKOTQQLQDADAVANDIKENIAIVERNTLLQALEBELRAVVBQTERSKLAEQELIER 1658  
QY 543 ---QINIPSSNDDCKRAEEVVSFIEPQEKEMEEFYEE-----REML 581  
DB 1659 SERVOQLHSQNTSLINQKKMDADLSQLQTEVEEAVOECRNABEAKAKAITDAAMAEEL 1718  
QY 582 IKDQEEK--MEDKKRHHETIFDLKEKFDALQOLMYKHG 619  
DB 1719 KKEQDTSAHLERKKNNEQTIKDLQHRLEDA--EQIALKGG 1757

Search completed: April 6, 2004, 19:36:00  
Job time : 26 secs

**This Page Blank (uspto)**

GenCode version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:34:56 ; Search time 46 Seconds  
(without alignments)  
3568.388 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 3313  
Sequence: 1 MESSRAGPMSEKKNVQGGYRP.....EPDEALEQIMYKHLNEDD 625

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/prodata/1/pubppa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/1/pubppa/US06\_NEW\_PUB.pep:\*  
3: /cgn2\_6/prodata/1/pubppa/US06\_PUBCOMB.pep:\*  
4: /cgn2\_6/prodata/1/pubppa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/prodata/1/pubppa/US07\_PUBCOMB.pep:\*  
6: /cgn2\_6/prodata/1/pubppa/US08\_NEW\_PUB.pep:\*  
7: /cgn2\_6/prodata/1/pubppa/US08\_PUBCOMB.pep:\*  
8: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep:\*  
9: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/prodata/1/pubppa/US09\_PUBCOMB.pep:\*  
13: /cgn2\_6/prodata/1/pubppa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/prodata/1/pubppa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/prodata/1/pubppa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/prodata/1/pubppa/US60\_NEW\_PUB.pep:\*  
17: /cgn2\_6/prodata/1/pubppa/US60\_PUBCOMB.pep:\*  
18: /cgn2\_6/prodata/1/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1428.5	43.1	611	US-10-425-114-58390	Sequence 58390, A
2	745.5	22.5	324	US-10-424-599-234018	Sequence 234018, A
3	721	21.8	223	US-10-424-599-234019	Sequence 234019, A
4	297	9.0	188	US-10-424-599-242612	Sequence 242612, A
5	258	7.8	219	US-10-424-599-178379	Sequence 178379, A
6	199.5	6.0	227	US-10-425-114-68938	Sequence 68938, A
7	199.5	6.0	414	US-10-425-114-63938	Sequence 63938, A
8	199.5	6.0	448	US-10-425-114-54044	Sequence 54044, A
9	182	5.5	1972	US-09-839-479-21	Sequence 21, Appl
10	178	5.3	1203	US-10-097-340-43	Sequence 43, Appl
11	175	5.3	1879	US-10-296-115-1265	Sequence 1265, Appl
12	173.5	5.2	1969	US-09-839-479-71	Sequence 71, Appl
13	172.5	5.2	593	US-10-363-616-295	Sequence 295, Appl
14	172.5	5.2	1156	US-10-369-493-43	Sequence 43, Appl
15	172	5.2	892	US-10-205-219-102	Sequence 102, Appl

16	170.5	5.1	676	US-10-256-250-16	Sequence 16, Appl
17	170.5	5.1	725	US-10-257-377-1	Sequence 1, Appl
18	170.5	5.1	725	US-10-256-250-14	Sequence 14, Appl
19	170.5	5.1	725	US-10-023-634-78	Sequence 78, Appl
20	170.5	5.1	725	US-10-023-634-79	Sequence 79, Appl
21	170.5	5.1	725	US-10-205-647A-4	Sequence 4, Appl
22	168	5.1	1043	US-10-310-154-449	Sequence 449, Appl
23	167.5	5.1	3899	US-10-171-311-4	Sequence 8, Appl
24	167.5	5.1	3917	US-10-171-311-8	Sequence 3425, A
25	166	5.0	345	US-10-029-386-34255	Sequence 2, Appl
26	166	5.0	1162	US-09-894-273-2	Sequence 2, Appl
27	166	5.0	1162	US-10-294-804-2	Sequence 2, Appl
28	165	5.0	724	US-10-023-634-80	Sequence 80, Appl
29	164	5.0	1790	US-10-369-493-1586	Sequence 1586, A
30	163.5	4.9	308	US-10-425-114-64137	Sequence 64137, A
31	163.5	4.9	498	US-10-104-047-3191	Sequence 3191, Appl
32	163	4.9	287	US-10-029-386-33041	Sequence 33041, A
33	162.5	4.9	709	US-10-256-250-15	Sequence 15, Appl
34	162.5	4.9	709	US-10-023-634-14	Sequence 14, Appl
35	162	4.9	1992	US-10-369-493-6527	Sequence 6527, Appl
36	161.5	4.9	1875	US-10-108-260A-4409	Sequence 4409, A
37	161.5	4.9	1875	US-10-369-493-22285	Sequence 22285, A
38	161.5	4.9	3878	US-10-080-608A-11	Sequence 11, Appl
39	161.5	4.9	3907	US-10-171-311-2	Sequence 2, Appl
40	161.5	4.9	3911	US-10-370-685-100	Sequence 100, Appl
41	161.5	4.9	3925	US-10-171-311-6	Sequence 6, Appl
42	161	4.9	650	US-10-104-047-3636	Sequence 3636, Appl
43	161	4.9	725	US-09-978-309A-47	Sequence 47, Appl
44	161	4.9	1837	US-10-369-493-22734	Sequence 22734, A
45	160.5	4.8	621	US-10-316-253-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1  
US-10-425-114-58390  
; Sequence 58390, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 58390  
; LENGTH: 611  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3079-030-A8\_F11.pep  
; US-10-425-114-58390  
  
Query Match 43.1%; Score 1428.5; DB 12; Length 611;  
Best Local Similarity 47.2%; Pred. No. 2.4e-99;  
Matches 294; Conservative 99; Mismatches 159; Indels 71; Gaps 15;  
  
QY 33 RLASGDDGGE-----WEVISKKNKKNKPGNTSGKTW-----VSQNSNPPRA 73  
DB 13 RSARSWSGSGDRRGCGPSPSHSGMETWKKSK-KPQAGARQWAPWSSTVTENTARP-A 70  
QY 74 WGGQQCGRGSSNVS---GRANNVSGRNGNGRGIGQANISGRGALSKYDNNFAPF-----126  
DB 71 WGGSGSHPSGTSWAQAPDHGATRNPR-----PSSQTS 105  
QY 127 -PVSRRPLRGGMWQARGSGAGHTAVOE-----PPDVED-DVDNASEENDSDALDSD 177



```

Db      106 RPLAFLPANGWQMSRPRSGSEVKKDDAPFSGSVSEVENVGNNTSDDDDD--DDL 163
Qy      178 DDDLADDDVDSVQSHSGSRKKNKFKPFGSLDLSLSTQINPQOROMHCPACONGPRA 237
Db      164 SDDI--SDDVDSDASEKSFETRKNKWFKEFEVLNTLSLQINPQOROMHCPACXNGPQA 222
Qy      238 IDMY-NLHPLLAHARTKGARVYKLHRELAEVLEKDLQMGASVTPCGEITYGQKGLGEDE 296
Db      223 IDMYKLOPLVSHARTKGTSTRVYKLHRELALALEEELSRKTSVLPBGEOPGKKGL--QSS 281
Qy      297 KQYEIVMPVMIIMNTRLDKDDNDKWLGMGNQELLEYFDKYEALRAHSHYSGPGRHGMV 356
Db      282 TBREIWPMPVIMVMTFLEKEDDDKWKMGQNGELLDYFGEYASKARHAYGSGRHGMV 341
Qy      357 LMFESSATGYLEAERLHRELAEVLEKDLRIAMGQKRSMP--SGGVQLYGFLATKDDLDLFPQ 415
Db      342 LTFESSAVGMEAEERLHKEFVNOGTRNSWHLRKVFPVPGKQLYGLFANKEDEAFNK 401
Qy      416 HSQKTRFLKFKELSYOEMVVKELROI SEDNQNLNFPKLSKONKAKVLEESLEINSEK 475
Db      402 HCHGSRKLTKEKMSYIEMVVIQKMS EDNQNLNFKKMKVTKQSKAVESLGVYTQK 461
Qy      476 LRRTAEDNRIVRQRTYQMOHEONREMDAHDREFMDSIKQIHERDAKEENFEMLOQOERA 535
Db      462 LRTEIENIFVRSKAKKEMVEEEMKSGEIRFHGLIEDIHKATEDKQFELQLQERS 521
Qy      536 KVVGOQOQINPSSNDCKRAEVSFIEFOKEMEFPVEEREMLIKQOKMEDMKR 595
Db      522 KA---RRFVDSGTMDCKRDLRKYYOKFTDCQYDAVEFEVDEDLIKVIEDKGLKKE 578
Qy      596 HHEIFDLEKEFDEALEQMLYK 618
Db      579 YMDLELEKEFPAALTGLMEK 601

```

```

RESULT 2
US-10-424-599-234018
; Sequence 234018, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234018
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1.pcp
US-10-424-599-234018

```

```

Query Match      22.5%; Score 745.5; DB 12; Length 324;
Best Local Similarity 31.1%; Pred. No. 4.3e-48;
Matches 162; Conservative 68; Mismatches 84; Indels 207; Gaps 6;

Qy      109 GGRALSRKYDNNFVAPPVSRPPLGGWNNQARGSAQHTAVQER--PDVEED--VDNA 164
Db      1 GGPQLTSGESNYVTNPNVIRPPLHGMWQSRPAIOSNVREDSPELDQKNYGVDD 60
Qy      165 SEENDSDALDDDDDLASDDVDVDSQSHSGSRKKNKFKFGSLDLSLSTQINPQOR 224
Db      61 GEEB-ESDLEDTDDDDSDASQSHSTRKSKKFKFELLDGLTVEQINEPER 119
Qy      225 QNHCPACONGPAIDMY-NLHPLLAHARTKGARVYKLHRELAEVLEKDLQMGASVTPCG 283

```

```

Db      120 QNHCPACONGPAIDMYNGLQPLVTHAKTSGRKVYKIHRELAILDEELRRRGSVTPPA 179
Qy      284 EITYGQKGLGEDEKQYEIWPMPVIMNTRLDKDDNDKWLGMGNQELLEYFPKYEALRAR 343
Db      180 EVFGKNGE----- 188
Qy      344 HSYGPOGHGMSVLMFESSATGYLEAERLHRELAEVLEKDLRIAMGQKRSMPSGGVQLYGF 403
Db      189 ----- 188
Qy      404 LATKDDLDLFNQSHQKTRFLKFKELSYOEMVVKELROI SEDNQNLNFPKLSKONKAK 463
Db      189 ----- 188
Qy      464 VLEESLEINSEKLRRTAEDNRIVRQRTYQMOHEONREMDAHDREFMDSIKQIHERDAKE 523
Db      189 -----KVQOKENKEEMSMQEEFPFQRTIHDSRAKE 221
Qy      524 ENFEMLOQOERAKVVGQOQONINPSSNDCKRAEVSFIEFOKEMEFPVEEREMLIK 583
Db      222 EEFERNQOKEKRV---KRSSTSPINEEGRVKVDYLKVFQFKEMENFVAABEKJLQ 278
Qy      584 DQKKNEDMKKSHHEIFDLEKEFDEALEQMLYKGLHNEED 624
Db      279 AHKDVNDAMTRRRHMEKVOLEERFNEBLAKLMEKXSLSHPE 319

```

```

RESULT 3
US-10-424-599-234019
; Sequence 234019, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234019
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(223)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53343C.1.pcp
US-10-424-599-234019

```

```

Query Match      21.8%; Score 721; DB 12; Length 223;
Best Local Similarity 60.1%; Pred. No. 1.9e-46;
Matches 131; Conservative 41; Mismatches 40; Indels 6; Gaps 4;

Qy      109 GGRALSRKYDNNFVAPPVSRPPLGGWNNQARGSAQHTAVQER--PDVEED--VDNA 164
Db      1 GGPQLTSGESNYVTNPNVIRPPLHGMWQSRPAIOSNVREDSPELDQKNYGVDD 60
Qy      165 SEENDSDALDDDDDLASDDVDVDSQSHSGSRKKNKFKFGSLDLSLSTQINPQOR 224
Db      61 GEEB-ESDLEDTDDDDSDASQSHSTRKSKKFKFELLDGLTVEQINEPER 119
Qy      225 QNHCPACONGPAIDMY-NLHPLLAHARTKGARVYKLHRELAEVLEKDLQMGASVTPCG 283
Db      120 QNHCPACONGPAIDMYNGLQPLVTHAKTSGRKVYKIHRELAILDEELRRRGSVTPPA 179
Qy      284 EITYGQKGLGEDEKQYEIWPMPVIMNTRLDKDDNDK 321
Db      180 EVFGKNGE----- 217

```



```

? RESULT 8
? US-10-425-114-54044
? Sequence 54044, Application US/10425114
? Publication No. US20040034888A1
? GENERAL INFORMATION:
? APPLICANT: Liu, Jingdong
? APPLICANT: Zhou, Yihua
? APPLICANT: Kovalic, David K.
? APPLICANT: Screen, Steven E
? APPLICANT: Tabaska, Jack E
? APPLICANT: Cao, Yongwei
? TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
? TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
? FILE REFERENCE: 38-21(53313)B
? CURRENT APPLICATION NUMBER: US/10/425,114
? CURRENT FILING DATE: 2003-04-28
? NUMBER OF SEQ ID NOS: 73128
? SEQ ID NO 54044
? LENGTH: 448
? TYPE: PRT
? ORGANISM: Zea mays

```

Query Match	5.54;	Score 182;	DB 12;	Length 1972;
Best Local Similarity	20.84;	Prod. No. 0.0018;		
Matches 113;	Conservative 93;	Mismatches 190;	Indels 158;	Gaps 20;

  

QY	140	QARGSSAGHTAVOEFPEVEDVDNASEENDSPALDDSDDLAS-----	183
Db	394	QPRGTDSDIPSSKSESDNEDEDEDEDEDEDEDESDSGSESDNSEPTGSEEE	453
QY	184	--DVDSDVSQKSHSGSRKONKMFKEFGSL--DSLSTEQINEQRMHCAPACONGGALD	239
Db	454	DDDDDDQDESDSDTEGEGKTSWKLNTKTTSSVKSPEMSTGHTSTRNLTHAKAPSDAPALC	513
QY	240	WYNLHPLL-----AHARTKAGRVYLHRELAEVLKED-----LQMR	275
Db	514	SSQSQALFGLTSSSTLTSSPHSGTSKRRRVDERELRIPLEYQWORETRIRNFGRLQGE	573
QY	276	GASVLPCEGIYQGM-----KGLGEDEKD-----YEIVV----	303
Db	574	VAYVAPCGCKLQYEPVITKYLSRNGLIMTISDINSFSFAKTRVGEDFYEARQDQPMQWCLL	633
QY	304	-----PPVYIINNTLTDKDDNDKMLGWNQOELLEFDYK--EALRRARHAYGPGCHGWS	355

Db 634 KEDVIRIRAMEGRGRPPNP-----DROARESESMRRRKGPRPVNGA 679  
 QY 356 VLMFESSA--TGYLEARHLELAEMGLDIRIAMGOKRSMPSGGVROLGYFLATKODLDIF 413  
 Db 680 EFLDNADALIKRLQOEIARQAQIKLRLKLOKODARVAKEAKQOALMAEE----- 734  
 QY 414 NQHSQKTELKELK-SYQEMVVKELROI SEDN---OOLNFKNLSKONKIAVLEES 468  
 Db 735 -----KRQKQKIKIMKQOEKIKRIQOI RMEKELAAQOILEAKKKKKKEAANAAILLEAE 788  
 QY 469 LEIMSEKLRLTADNRIVQRTYMOHEONREMDADRFPMSIKOIHERRDAKEENFEM 528  
 Db 789 KRIKEKEMER-----QOAVLKKH-QSERERRQH-----MMLKAMEARKKAEK--ER 833  
 QY 529 LQOGEAKVVGQOQONINSSNDCKRAEVSSTIEFOKEMEVEVEREMLIKQOEK 588  
 Db 834 LKOEK-----DEKRLN-----KERKLEQRLELEM-AKELKKP 866  
 QY 589 MEDM 592  
 Db 867 NEDM 870

RESULT 10  
 US-10-097-340-43  
 ; Sequence 43, Application US/10097340  
 ; Publication No. US20030087250A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: John MONAHAN  
 ; APPLICANT: Manjula GANNAVARAPU  
 ; APPLICANT: Sebastian HOERSTCH  
 ; APPLICANT: Shubhangi KHAMATKAR  
 ; APPLICANT: Steve G. KOVATS  
 ; APPLICANT: Rachel E. MEYERS  
 ; APPLICANT: Michael MORRISSEY  
 ; APPLICANT: Peter OLANDR  
 ; APPLICANT: Ami SEN  
 ; APPLICANT: Peter VEIBY  
 ; APPLICANT: Gordon C. BAST, Jr.  
 ; APPLICANT: Karen LU  
 ; APPLICANT: Rosemarie SCHMANDT  
 ; APPLICANT: Xumei ZHAO  
 ; APPLICANT: Karen GLATT  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
 ; FILE REFERENCE: MRI-030  
 ; CURRENT APPLICATION NUMBER: US/10/097,340  
 ; PRIOR FILING DATE: 2002-03-14  
 ; PRIOR APPLICATION NUMBER: 60/276,025  
 ; PRIOR FILING DATE: 2001-03-14  
 ; PRIOR APPLICATION NUMBER: 60/325,149  
 ; PRIOR FILING DATE: 2001-09-26  
 ; PRIOR APPLICATION NUMBER: 60/276,026  
 ; PRIOR FILING DATE: 2001-03-14  
 ; PRIOR APPLICATION NUMBER: 60/324,967  
 ; PRIOR FILING DATE: 2001/09/26  
 ; PRIOR APPLICATION NUMBER: 60/311,732  
 ; PRIOR FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: 60/325,102  
 ; PRIOR FILING DATE: 2001-09-26  
 ; PRIOR APPLICATION NUMBER: 60/323,580  
 ; PRIOR FILING DATE: 2001-09-19  
 ; NUMBER OF SEQ ID NOS: 363  
 ; SOFTWARE: FaastSeq for Windows Version 4.0  
 ; SEQ ID NO 43  
 ; LENGTH: 1203  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-097-340-43  
 Query Match 5.4%; Score 178; DB 14; Length 1203;  
 Best Local Similarity 19.0%; Pred. No. 0.00019;

Matches 173; Conservative 121; Mismatches 284; Indels 332; Gaps 40;  
 QY 15 OGQYRPE-----VEOLVQGLAGTR--LASODDGGEMEYISKONKNGKSGKTM 63  
 Db 37 RQGRPRKAKARASTYGVAVRVGIRAGPVLVNSGKGSDSGVQIKGANDG--ASGA-- 93  
 QY 64 VSQNSNPRAWGQOQOGRGNSVNGRNNVSGRNGNGRGI-----QANISGRGALSRYKD 119  
 Db 94 LSSDLELPNPFYSQYKGFAPAPQSSSTDEEPAGYNNKGLRSHSQASLSLGPVDSNRS 153  
 QY 120 NNFV-APPVSRP-----PLE-----GKNMWARGSSAHTVAQEPDYE 158  
 Db 154 NSMTLEIAPVYASPGSTIDTAPLSSVDSLINKEDSOLG--QARGTGRRTM--LPBQ 208  
 QY 159 DDVDNASEENSDALDDDD-----LASDDYDSVSGKSHGRKONKMPKFFBSLDS 213  
 Db 209 RRRKSLDSRLPRDTPFEEROSTNHWTSSTYDHNVTGSKOPASQON-----LSP 259  
 QY 214 LS-----IEQINEPOROWHCPACON-----GPAIDWY--NLH 244  
 Db 260 LSGFSRSRQTQDWVLGFEPRRSADPTMLQFKSTPDLRLDQGEAAPGSDVHMKATY 319  
 QY 245 PLLAHARTKGARRVKILHRELAYLEKLOMRGASVTPGE--ITYQ-----WIG 291  
 Db 320 GILREGSSESETSVR--RKVSIVLEK--MOPLVWVSSGSTRVAVAGSELTRKVELORK 374  
 QY 292 LGEDKDYEIWPPWYIINWTLRLDKDNDKMLGMSGNELLE-----Y 333  
 Db 375 LDEEVKKRKLPSQVGLEROLEBEKTEBCSRL--QELLERKKGAAQOSKELONMKEL 430  
 QY 334 PDKYEALR-----ARHSYQPO----- 349  
 Db 431 LQGBDLRHGLTQYMWELONKIKHVGPEPAKEVILLKOLLETRELLEVLBGQRYEQL 490  
 QY 350 --GHRGSVYL--MFESSAT-----GYLEAR-- 371  
 Db 491 RLREBELTALKALKEEVASRDQEVHVRQYORDTEQRRSMQDATODHAVALLEARQK 550  
 QY 372 -----LHRELAEMGLDIRIAMGOKRSMPSGGVROLGYFLATKODL----- 410  
 Db 551 SALVRGLQRELETSSETGHW--OSMFQKNKEDL--RATYOELLQRLMEKEMEELG 604  
 QY 411 ---DIFNOH-----SOGKTR--LKFEKSYQEVVVKEL----- 438  
 Db 605 EKIEVIGRLBEQARASAGTROYEVYLKELLATQF-ELKELQAEKRSQVAGRHRRLE 663  
 QY 439 -----ROISEDNOQ--LNYFNKLSKONKIAVLEESLIMEK----- 475  
 Db 664 KOLAVLRYVADRGRELEBQNLQLOKTLQDRDCEBASAKVVALEAATVLCGRRAVET 723  
 QY 476 -LRTTADNRIVQRTYMOHEONREEM-----DAHDFPMSIKOIHERRDAKEENFEM 528  
 Db 724 TLRETOEENDEFFRRRLTGLEQQLKETRGVLVDGEAVALRLKLOLEAKQOLEBALNA 783  
 QY 529 LQOGE-----RAKY-----VGQOQONINPSNDCKRAEVSSTIEFOR- 568  
 Db 784 SQEBEGSLAAARGALEARLEBAQRLAGQOQOTLNLRLBEGKREVLARKKALEBQ 843  
 QY 569 -----KEMEEFVEEREMLIKQOEKMEKMKHHEITPDEKEF-DEALEQLM 615  
 Db 844 KRLDRTVRLNKLEKIGEDSKQALQOQOLEDEYKARREVADAGQARDWASEAK 903  
 QY 616 YKHGLHNEED 625  
 Db 904 TSGGLSRLDD 913

RESULT 11  
 US-10-296-115-1265  
 ; Sequence 1265, Application US/10296115  
 ; Publication No. US20040053248A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq Inc

```

1  TITLE OF INVENTION:  NO.  US20040053248A1el Nucleic Acids and Polypeptides
2  FILE REFERENCE:  784PCT
3  CURRENT APPLICATION NUMBER:  US/10/296,115
4  CURRENT FILING DATE:  2002-11-18
5  PRIOR APPLICATION NUMBER:  US09/488,725
6  PRIOR FILING DATE:  2000-01-21
7  PRIOR APPLICATION NUMBER:  US09/552,317
8  PRIOR FILING DATE:  2000-04-25
9  NUMBER OF SEQ ID NOS:  1478
10  SEQ ID NO 1265
11  LENGTH:  1879
12  TYPE:  PRT
13  ORGANISM:  Homo sapiens
14  FEATURE:
15  NAME/KEY:  misc_feature
16  LOCATION:  (1)..(1879)
17  OTHER INFORMATION:  Xaa = any amino acid or other as shown in Table 3
18  US-10-296-115-1265

```

Query Match	5.3%;	Score 175;	DB 12;	Length 1879;
Best Local Similarity	20.2%;	Pred. No. 0.00057;		
Matches 116; Conservative	94;	Mismatches 194;	Indels 170;	Gaps 24;

Qy	158	EDVDVNA	SEEE-----	ENDSDALDSDDD	LASDDYDSDVSQKSHSGSKGNKFWFK	206
Dd	1297	QKEIENLTQY	VEEKAAAYDKLEKTKNRLQOEELDVLV-DLDNGRLVSNLEKKÖRK-FDQ	1354		
Qy	207	FFGSJDSLSIT	EOINPEPOQWCHCPACQNGPAILDWYNLHPILAAARTTKGARVYKHEILAE	266		
Dd	1355	ILAEKKNIS	SKVADRDR-----	VEAEKREKETALSIARLAE	1393	
Qy	267	VLEKOLMR	GA	SVIPCEI-----	YGQWKGLED-EKQYEIYWPWVJINMTRLD	315
Dd	1394	ALAEAEEL	ERTKMKLAEKMGPGSASKDVQÖ--ELSHDEKS	KRALGDRELEMTÖLE	1451	
Qy	316	KDDNDKWL	CMQNOEL-----	LEY-----	FDKYEARARSHYSGPOHGRMSV	356
Dd	1452	E-----	LGRTIELASPRRDAKRLLENVMOA	PSRPAFER--DLOAATTEQNESSRHHTQR	1501	
Qy	357	LMPESATG	VLEAEERLHRELAEMGLDRIAMQKSM-----	FFGG--VROLYGFLLA	405	
Dd	1502	QIHEYE	TE--LEDERKÖRALMAAAKIKIGMDPVTRLDXADSAIKGRGKAIKÖRLQÖA	1559		
Qy	406	TKODL-----	DIF--NOHSÖQTKLTKELKS	YQEMVYKELRQISENOOLNY	450	
Dd	1560	QMDQFQ	ELDAPASRDEIFATAKEXENKAKSLIEDLMOÖLE----	DILAAEBGRKQADL	1615	
Qy	451	FKAKLS-----	KÖNGHAY--LEESTL-----	EIMSEKLR--	478	
Dd	1616	EKEELAE	LASSLSGRNALODEKRLAEKAIQÖLEEBLEEQNGEMAMSDRYAKATQÖAQ	1675		
Qy	479	-----	TAEDNRIVRÖRTQMOHEQÖNEEMDAHDFRPFMDSTIQIHERDAX-----	EE	524	
Dd	1676	LISNELAT	ERSHQAQNGESARQÖLEKÖNKELSKHEWEGAVKSKFSTTALAEAKIQAÖLE	1735		
Qy	525	NFEMLQÖ	ERAKVVGQÖQÖQININSNDCKRAEEVSSFIEQÖKEMEERFVEERZ-----	579		
Dd	1736	QVEÖBARE	KQATKSLKÖKD-----	KKKEILLQVEDERKMAEÖYKQÖAKGNAR	1785	
Qy	580	-MLI	DOEKMEDKGRHHEEIFDLEKEFPDEALE	612		
Dd	1786	VKÖLRQÖ	LEAEBSÖQRINAMRRLQÖLEDEATE	1819		

RESULT 12  
US-09-839-479-71  
Sequence 71, Application US/09839479  
Publication No. US20020039719A1  
GENERAL INFORMATION:  
APPLICANT: Jones, Michael H.  
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR  
FILE REFERENCE: 06501-042002  
CURRENT APPLICATION NUMBER: US/09/839,479

```

; CURRENT FILING DATE: 2001-04-20
;
; PRIOR APPLICATION NUMBER: US 09/418,710
;
; PRIOR FILING DATE: 1999-10-15
;
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
;
; PRIOR FILING DATE: 1998-04-17
;
; PRIOR APPLICATION NUMBER: JP 9/310027
;
; PRIOR FILING DATE: 1997-10-24
;
; PRIOR APPLICATION NUMBER: JP 9/116570
;
; PRIOR FILING DATE: 1997-04-18
;
; NUMBER OF SEQ ID NOS: 72
;
; SOFTWARE: FaSTSeq for Windows Version 4.0
;
; SEQ ID NO 71
;
; LENGTH: 1969
;
; TYPE: prt
;
; ORGANISM: Homo sapiens
;
US-09-439-479-71

```

Query Match	5.2%;	Score 173.5;	DB 12;	Length 1969;
Best Local Similarity	20.1%;	Pred. No. 0.00079;		
Matches 128;	Conservative 94;	Mismatches 213;	Indels 203;	Gaps 25;

QY	47	ISKANAKPNTSOKTIVWSQNSNPRAMGGQOQSGNSGGRGNVSGRNGNG-----	100
Db	341	ISTSGRRTPG--OTPMVPSAPIL--HSQKEKAVSNVAVPTQHSHPAKSLVE	393
QY	101	--RGIOANISGRGALSHKYDNNFVAPPPVSRPFLBEGMMKMGAGSAQHTAVOEPDVE	158
Db	394	QFRGTDSI-----PSSK-----DBSDSDEDEEDDEE	421
QY	159	DVDVNASHEENDSDALDS-----DDLASDDYDSDVSQKSHGSRKQNMWF	205
Db	422	EDEDEDDSDSDSGBESDSSNSBSTDGSEBDEDDDKODESDSD---TEGKTSMLNK	477
QY	206	KFFGSDSISIQINEPQOMHCPCQNGPAGIDMYNHLPL-----AHARTG	254
Db	478	TTSKSPMSLTHGSHPPNLIHAKPGAPALCBESQSPALGTSSTLTSPSHGTSK	537
QY	255	AREVTLHRELAETLEKD-----LQMGASVIPCCEITGOM-----KGL	292
Db	538	RRAVTERELRLPLEYGMQRETRIRNFGRLQGEVAYYAPCCKLRQYPEVYKYSRGI	597
QY	293	GEDEKD-----YEIYW-----PMVYIMNTRLTDKDDNDKWLG	324
Db	598	WDISRDFSSAKIRVGDYEADGQEWQWCLLKEEDVYPRIRAMEGRGRPPN---	653
QY	325	MGNQELLEYPDKY--EALPARHSYPOGHRGMSVLMFESSA--TGYLAEARLHRELAEM	379
Db	654	-----DROARARESEMRRRRKRPPVNGAEFLDNADAKLRLKQAQEA--ROAQI	702
QY	380	GLDIRAWGKRSMPGSGVROLVGLATKODDIFNCHSQGKTRLXELKSY--QEWVYKEL	438
Db	703	KLRLKLOQOBARAVAKXAKQAQALMAAE-----KROKQOKIHKQOEKIKI	751
QY	439	ROISEDN---QOLNYFKNLSKONGKAKYLEISLEIMSEKLRTRTENDNRIVORTMOH	494
Db	752	QOIRWEKELRAQOFTLEAKKKKKKEAANAANKLLEKRIKEREMR-----BOQAVLTK	802
QY	495	EONREEMDAHRFWDSIKQIHERDAKENFENLQOQEBAKVVGQOQOINSSNDCH	554
Db	803	QRRERRRROH---MLMKAMEARKKKEK--EKLKEKR-----DEKLN-----	842
QY	555	KRAEVSSFIEQKEMEPEVEEREMLIKQOEKQMEDM	592
Db	843	-----KERKLRORLLEEM--AKELKKPREDM	867

RESULT 13  
 US-10-363-616-295  
 ; Sequence 295, Application US/10363616  
 ; Publication No. US20040044181A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hyseq, Inc  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

FILE REFERENCE: 21272-113 (793)  
 CURRENT APPLICATION NUMBER: US/10/363,616  
 CURRENT FILING DATE: 2003-03-03  
 PRIOR APPLICATION NUMBER: 09/654,935  
 PRIOR FILING DATE: 2000-09-01  
 NUMBER OF SEQ ID NOS: 490  
 SEQ ID NO 295  
 LENGTH: 593  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-363-616-295

Query Match 5.2%; Score 172.5; DB 12; Length 593;  
 Best Local Similarity 19.0%; Pred. No. 0.00019;  
 Matches 110; Conservative 100; Mismatches 209; Indels 161; Gaps 23;

162 DNASEENSDALDDSDLDASDDYSDVSQKSHGSRKQNK-----WFK 205  
 52 DSTAE-----ADELDGEIDLSCTDYTEAYQVRYVGFQIHTKDAVYTLAMTSGIRRWIE 108  
 206 KFRGSLDSLSIEDI-----NEPQRQWCPACONGP-----235  
 109 ALAKTRPFTSAPDYTKLSDNKNEMALHSYSTQKGPLKAGORAGSEVISRKGPRKADQOR 168  
 236 GAIDWYNLHPL-----AHARTKGRARYKHLRELAELVKDLOMKG-----AS 278  
 169 QALDYVELSPLTQASPPARTPARTPRLAKQ-----EELERDLQSRSEERRKMFATDSRT 226  
 279 VIFQGEIYGQWKLGEDEKQEIYVWPVVIIMTRLDKXNDK-----322  
 227 EVPAGE--GPRRGIG-----APLTEDQQRNLSSEIEKTKQOELEKPLRENKRVPL 274  
 323 LGMNG-----ELLEVPDKYELRPAR-HSYGPOGRKMGVLMFESSA---TG 365  
 275 TALLNOSRGERRPPSDGHEALE--KEVQALRAQLKELWRLQSGAPALRSQEDGHIIPG 332  
 366 YLEAERLHRELAEMG-----LDRIAWGQKRSWFSGGVROLYGFATKODDIFNOHSG 419  
 333 YIQGEACERSLAMESHQVMELOKHREKOR-----LQOEKWLAEFTMAT 383  
 420 KTRLEKELKSYQEMVVKELRQISEDNQOLNYFNKLSKQNK-HAKVLEESLEIMSEKL-R 477  
 384 ASAIEAMKRAYQELSRLESK-----TRSLQCGPDGLKQHSQVEALKRELQVLSQYSE 439  
 478 RTMEDRIYQRTKMGH-----EONREMDAHRFPMDSTKQIHERDAKEENFEMIQOQ 532  
 440 KCEIGALMKAQAEERHETLRCCOEGEELRN-----OELHGR--LSEETDOLRGF 489  
 533 ERAKVVQOOQONINPSSNDCCRRAEVSSEFIEQEKMEEFVEERMLIKDOE-----586  
 490 IASQMGNGGGRNENSSCELEVLARKENELQYKKEVGLADELOMGKDKRFTSGKY 549  
 587 ---KMEKMKRHHHEIFDLKEFDEALEQMLYKGLHN 622  
 550 QDYVELSHIKTSSEREIEQLKEHLRLAMALQEKESMRN 589

RESULT 14  
 US-10-369-493-43  
 Sequence 43, Application US/10369493  
 Publication No. US20030233675A1  
 GENERAL INFORMATION:  
 APPLICANT: Cao, Yongwei  
 APPLICANT: Hinkle, Gregory J.  
 APPLICANT: Slater, Steven C.  
 APPLICANT: Goldman, Barry S.  
 APPLICANT: Chen, Xianfang  
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 FILE REFERENCE: 38-10(52052)B  
 CURRENT APPLICATION NUMBER: US/10/369,493  
 CURRENT FILING DATE: 2003-02-28  
 PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21  
 NUMBER OF SEQ ID NOS: 47374  
 SEQ ID NO 43  
 LENGTH: 1156  
 TYPE: PRT  
 ORGANISM: Aquifex aeolicus  
 US-10-369-493-43

Query Match 5.2%; Score 172.5; DB 15; Length 1156;  
 Best Local Similarity 19.9%; Pred. No. 0.00047;  
 Matches 82; Conservative 76; Mismatches 139; Indels 115; Gaps 15;

268 LKEDLQMGASVIPCSEI-----YQGMKGI-----GEDEKQYELWPPMYI 308  
 555 LAKRMILGRFSFPLNRAVVEERPLYPRTKGVADPAVNLVEXDPFEKVKVFPDITL 614  
 309 IMNTRLDKXNDKMLGNGQELL-----EYFDKELRAHSHSGPOGRKMSVLMFESSAT 364  
 615 VENF-----ESAKVIGIGYRWVTLLEGELFEKSGVI-----TGAV-----KPS 653  
 365 GYLEAERLHRELAEMGLDRIAWGQKRSWFSGGVROLYGFATKODDIFNOHSGKTRPLK 424  
 654 GEINKRYEEBELQRLAESEKLNESITIOKTRIRNLISEKTL-----LK 701  
 425 FELKSYQEMVVKELRQISEDNQOLNYFNKLSKQNKHAKVLEESLEIMSEKLRFAED-- 482  
 702 VSRKIEBELSEBELGYEEK-----FKKLENSKEYLKLLEKTLNVDDKLELABEIE 755  
 483 ---NRIVORTKQHEONREMDAHRFPMDSTKQIHE-----RDAKEENFEM 528  
 756 YEEKLNILKLEKGDIKRHYSGVEKREKREYSKRVQSELEKSLNEIERELNKTYEL 815  
 529 ---LQOEAKVQOOQONINPSSNDCCRRAEVSSEFIEQEKMEEFVEERMLI 582  
 816 EYLEKEIEQEKERR-----EYLTERRISLKEEINLILFEKKTQOE-VKEAIVKV 864  
 583 KQOEKMEKMK-----RHHEIFDLKE--FDEALEQL 614  
 865 YDYIKKEBELKELILNLSKLGKLIKERELEKKEIKERKNLKVLEKIENTL 916

RESULT 15  
 US-10-205-219-102  
 Sequence 102, Application US/10205219  
 Publication No. US20030138803A1  
 GENERAL INFORMATION:  
 APPLICANT: Warner-Lambert Company  
 APPLICANT: Lee, Kevin  
 APPLICANT: Dixon, Alistair  
 APPLICANT: Brooksbank, Robert  
 APPLICANT: Pinnock, Robert  
 TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain  
 FILE REFERENCE: WL-A-018200  
 CURRENT APPLICATION NUMBER: US/10/205,219  
 CURRENT FILING DATE: 2002-07-24  
 PRIOR APPLICATION NUMBER: GB 0118354.0  
 PRIOR FILING DATE: 2001-07-27  
 NUMBER OF SEQ ID NOS: 197  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 102  
 LENGTH: 892  
 TYPE: PRT  
 ORGANISM: Rattus norvegicus  
 OTHER INFORMATION: Myosin heavy chain  
 US-10-205-219-102

Query Match 5.2%; Score 172; DB 14; Length 892;  
 Best Local Similarity 21.0%; Pred. No. 0.00036;  
 Matches 122; Conservative 98; Mismatches 186; Indels 174; Gaps 29;  
 141 ARGSAQHTAVQFP--DVEDDYDNNSEENSDALDSD--DDDLASDDYSDVSQKSHGS 197

```

Db      259 SRGKQAFITQIIEELKQLEEVYKAKSL--AHALQSSRHDCDLLREQYEEQPAKELQ 315
QY      198 RKQNK-----WPKKFGSLDLSLIEQINEPQRMHCPACONGPAIDWYNLHPLLAHA 250
      316 RAMSKANSEVAQWRTYX---ETDAIQTEB-----LEEA 346
Db      251 RTGARRVYLHRELAFLVEKDLQMRGASVIPCQEIYQWKGLGEDEKDYELVWPPWYIM 310
QY      347 KKLQQLQDAEHVAVN-----AKCASLEKTKQRLQNEVEDL-----MIDVE 390
Db      311 NTR-----LDKDNDRKMLGNQOELFEPDKYEALRHRHSYGPQGRGMSVLMF----- 359
QY      391 RTNACALADKKQKN-----FDKILAEWQKYEETTHAELEASQKSRSLSTELFKKNAY 445
Db      360 ---ESSATGYLEAEERLHRELAEMGLDRILAMGQKR---SMFSGVRLYGFLL-ATKQDL 410
QY      446 EESLDQLETLKREKNLQOEISDL-TEQLAEGGRITHELEKTIKQIEQEKSEILOALEEA 504
Db      411 DIFNQSQK-TRIKFEIKSYQEVNVELRQISEDNQOLNYPKNKLSKONKHAKVLEESL 469
QY      505 EASLEHEBGKILRIQLEINQVKEIDRKIAKDEIDQI-----KKN-HIRVESWQ 555
Db      470 EIMSEKLRRTAEDRRIYQRTKMOEQONREEMDAH--DRPFMDSIK-----OI 515
QY      556 STLDAEIR--SRNDALRIKKMEGDLNEMEIQLNHSNRMAAALRNYRNTQGIKDTQL 612
Db      516 H-----ERR---DAK--FENFEMLOQOERAKVVGQO----- 542
QY      613 HIDDALRGQEDLKEQLAMVERRANILQAEIEELRATLEQTERSKRIAEQELLDASRVQL 672
Db      543 ---QINPSNDDCKRAEVSFFIEQEKEMEVEEB-----REML 581
QY      673 LHTON---TSLINTKKLEETDISQI---QGEWEDIQGEARNAEKAKKAITDAMMAEEL 726
Db      582 IKDQEKK--MEDKKRHHHEITFDLEKEPDEALQOLMYKHG 619
QY      727 KKEQDTSALHERKKNLEQTVKDLQHRIDEA-EQALAKGG 765

```

Search completed: April 6, 2004, 19:40:48  
 Job time : 49 secs







QY	1921	ACCTGGGGGTCAAGAACCGCAGATAGTTTTCTGGAAGGTGTCGCAACCTGATAGGCTCTT	1980
Dp	1921	GCCTGGGGGTCAAGAACCGCAGATAGTTTTCTGGAAGGTGTCGCAACCTGATAGGCTCTT	1980
QY	1981	GCAACGAAAGCAAGATCTTGACATATTCATCAACACTCTCAAGGTTCTCTCCCCAAGA	2040
Dp	1981	GCAACGAAAGCAAGATCTTGACATATTCATCAACACTCTCAAGGTTCTCTCCCCAAGA	2040
QY	2041	AATTTGATATATGCTTTTAGTGTTCATGTGAATTAAAGTTTTGTTGTCCTCGTTTAA	2100
Dp	2041	AATTTGATATATGCTTTTAGTGTTCATGTGAATTAAAGTTTTGTTGTCCTCGTTTAA	2100
QY	2101	TGCAATCTGTTATGTATATATCTATGATTCATTAGGCAAAACAAGCTGAAATTTGCAAGTTG	2160
Dp	2101	TGCAATCTGTTATGTATATATCTATGATTCATTAGGCAAAACAAGCTGAAATTTGCAAGTTG	2160
QY	2161	AAATCATATCCAAAGAGATGGTGTGTAAAGAGCTGAGGCGAGATCTCTGAGAACATCAAGCAG	2220
Dp	2161	AAATCATATCCAAAGAGATGGTGTGTAAAGAGCTGAGGCGAGATCTCTGAGAACATCAAGCAG	2220
QY	2221	CTGAACCTATCTTTAAGAACAAAGCTCTCAAAAAGAAACAAGACGCCAAGGTCCTTAGAGAA	2280
Dp	2221	CTGAACCTATCTTTAAGAACAAAGCTCTCAAAAAGAAACAAGACGCCAAGGTCCTTAGAGAA	2280
QY	2281	TCTCTGGAAATTTATGAGCGGAGAGCTGCGCTGTGAAGCTGACAGAGATATTCGATCGTGA	2340
Dp	2281	TCTCTGGAAATTTATGAGCGGAGAGCTGCGCTGTGAAGCTGACAGAGATATTCGATCGTGA	2340
QY	2341	CAGAGAACTTAAGATGACGATGAACAGAAACAGGGAAGGATATGATTTTCTCTGAAATAT	2400
Dp	2341	CAGAGAACTTAAGATGACGATGAACAGAAACAGGGAAGGATATGATTTTCTCTGAAATAT	2400
QY	2401	CACAAACTTGAACATTTTGTATTTACTCTGATTCACATTTTGTGATATATATGTCACAA	2460
Dp	2401	CACAAACTTGAACATTTTGTATTTACTCTGATTCACATTTTGTGATATATATGTCACAA	2460
QY	2461	AAAACTGTGTGGTTTGAAGATGATCAACAAGATTTTTCATGATTCATCAATCAA	2520
Dp	2461	AAAACTGTGTGGTTTGAAGATGATCAACAAGATTTTTCATGATTCATCAATCAA	2520
QY	2521	CAGATCCATGAAAGAGAGACGCAAAAGAGGAAATTTGAGATGTCACAGACAGAGAA	2580
Dp	2521	CAGATCCATGAAAGAGAGACGCAAAAGAGGAAATTTGAGATGTCACAGACAGAGAA	2580
QY	2581	CGTGCAGAGTGTGTTGGCCACAGCAGCAGCAAACTTAATTCCTCTTAGCAATGACGATTC	2640
Dp	2581	CGTGCAGAGTGTGTTGGCCACAGCAGCAGCAAACTTAATTCCTCTTAGCAATGACGATTC	2640
QY	2641	CGAAAGAGGTATATGTACTTAACTTAACATAATTCCTCTGCGCTTTTGTTTTCAAACTTA	2700
Dp	2641	CGAAAGAGGTATATGTACTTAACTTAACTTAATTCCTCTGCGCTTTTGTTTTCAAACTTA	2700
QY	2701	AGAGTAATCTGAATATATTCGCGTTTTGATTTCTTTGCSAAGCTGAGGAAGTCAAGCTTC	2760
Dp	2701	AGAGTAATCTGAATATATTCGCGTTTTGATTTCTTTGCSAAGCTGAGGAAGTCAAGCTTC	2760
QY	2761	ATCGAGTTTCAAGAGAAAGAGATGAGAGTTTGTGAAAGAGAGGGAATGCTGTATAA	2820
Dp	2761	ATCGAGTTTCAAGAGAAAGAGATGAGAGTTTGTGAAAGAGAGGGAATGCTGTATAA	2820
QY	2821	GATCAAGAGAAAGATGGAAGCACTGAAGAAGAGGCATCACAGAGATATTTGATCTG	2880
Dp	2821	GATCAAGAGAAAGATGGAAGCACTGAAGAAGAGGCATCACAGAGATATTTGATCTG	2880
QY	2881	GAGAAAGAAATTTGATGAGGCTTTTGAACAAGCTCATGTATCAAGCATGGCTTCACAATGA	2940
Dp	2881	GAGAAAGAAATTTGATGAGGCTTTTGAACAAGCTCATGTATCAAGCATGGCTTCACAATGA	2940
QY	2941	GATGATTTAGACAATAAGTCTGTATACACAAGACTTAAGTTTCTTTGTTTGCTTTTG	3000
Dp	2941	GATGATTTAGACAATAAGTCTGTATACACAAGACTTAAGTTTCTTTGTTTGCTTTTG	3000
QY	3001	GTAATGTCGAAAGTATGAGATCTGAGACATTCATTTAATATCTAGCAAAATCTTAAGCA	3060

Accession	Sequence	Position
D8	GTATGTCGGAAAGTAGGAGATCTGAGAGACGTCCATTTTAAATACTAGGACCAAAATCTAAGGA	3060
OY	GATATATGATTAATATCTCCAAATTTTATGACGGATCTAAGSAGCATTAAGTCTT	3120
D8	GATTATGATTAATTAATCTCCCAATTTTATGACGGATCTAAGSAGCATTAAGTCTT	3120
OY	GTGACTAAAAACAAGTTTCCCTAGTATTTTGTTTTTTGGTAAAAATTTCAATGGAAG	3180
D8	GTGACTAAAAACAAGTTTCCCTAGTATTTTGTTTTTTGGTAAAAATTTCAATGGAAG	3180
OY	TTAGACATATTACCAAAAGTCAGAGTAATCACAGAATGCAAAATCATATGTTTT	3240
D8	TTAACAATATTACCAAAAGTCAGAGTAATCACAGAATGCAAAATCATATGTTTT	3240
OY	AGATTTTATATCTACAAAATATATGGTCCAAAT	3275
D8	AGATTTTATATCTACAAAATATATGGTCCAAAT	3275

RESULT 2	
AB025633	AB025633 81365 bp DNA linear PLN 27-DEC-2000
LOCUS	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:M0M1.
DEFINITION	AB025633 BA000015
ACCESSION	AB025633.2 GI:10178221
VERSION	
KEYWORDS	
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (sites)
REFERENCE	Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H. and Tsubae,S. Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones DNA Res. 7 (1), 31-63 (2000)
TITLE	
JOURNAL	DNA Res. 7 (1), 31-63 (2000)
MEDLINE	20181125
PubMed	10718197
REFERENCE	2 (bases 1 to 81365)
AUTHORS	Nakamura,Y.
TITLE	Direct Submission
JOURNAL	Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yinakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934) On Sep 15, 2000 this sequence version replaced gi:4589439. Address for correspondence: kao@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/and_graph.cgi?c=M0M1 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Graal (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Graal-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremlin.zool.iastate.edu/cgi-bin/bp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is K19M13 and the 3' clone is M0M1. Location/Qualifiers 1..81365 /organism="Arabidopsis thaliana" /mol_type="genomic DNA"

/strain="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="WQM1"  
/clone\_11b="Witcou P1"  
join(3184..3403,3474..3595,3681..3723,3777..3898,  
4042..4198,4445..4515,4601..4714,4888..4988,5300..5420)  
/note="gb|AAFO1580.1  
gene\_id:MOM1.24  
similar to unknown protein"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="BAA97251.1"  
/db\_xref="GI:8809710"  
/translation="MHVPGTSDVESPAHKVDETPPKHVOFLBPSIKTVVDAQNP  
VGSAPDPSSNSPLSPVEEPPSSSEGGDDPLPGIEDLOISGEYPGHELOACG  
YSLNGTSCNPFVWCHLEDGVSIVYDESSRETSNAKQPNVLTADVDVLYLAEV  
QPLDDNRNGKPKRKYRDLTVLIDVHTLLPSADCSVPEFMMFALICVPLISGG  
ACEDPMOSNIETKLTHTGASVKSLSAVGVDFIMEKATLSIREGYSIKCISDLTIAE  
KESASTVTIPFGQPAELV1IGSDGSHSLRADNSPDIGSRDEIVLRLFIKRM  
FFKALSFYHFRCSCHSRGNL"  
join(6317..6428,6508..6629,6714..6756,6841..6932,  
7121..7169,7396..7466,7550..7663,7785..7882,8112..8181)  
/note="gb|AAFO1580.1  
gene\_id:MOM1.23  
similar to unknown protein"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="BAA97250.1"  
/db\_xref="GI:8809709"  
/translation="MQGTSVSPALNOADEPPISTKVNDTKNIFEDDDAPL  
PALENIQISEPYGHELOACGSLNGTSCNFEWCHLEDGVSIVYDESSRETSNAKQ  
ADVGLCLAEVQPLDDNRNGKELVAFANDNKIACHPEMOSNIDKTLHTGASV  
SLAIFVHWEATLSIEREGYIKCNNDLTITEKFSATVAVIKPEKAEVLTIGSD  
GSHCLRVNDWEPDIRLCKEKKVRLVKVYVFEQKL"  
complement(join(8642..8705,8794..8863,8978..9387,  
9384..10347))  
/note="gene\_id:MOM1.22  
unknown protein"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="BAA97249.1"  
/db\_xref="GI:8809708"  
/translation="MSMWKSGSGMTAFDLKOKOGCLSEVEGDPFPVSTVNASF  
GVRLRLRNEPSEKSFSSVLPSPRPALTENDCQNGRCCCKRKPPTSLPNS  
HDLAFPTLKEMNSWADNLIIRDVLSTEDGFEMALFLKGVSSGKDEEPTSIKGY  
SSDRRESEYRFEKTVTSVMAARSTFEDAGKVLKENSQSSFLVNASNEKFPDI  
SELDSIIORLOSIPIEPEWEDDYLISHKDALKVMKASNSHRAQNAQORVHASA  
KQSDKAREMDLAEKLENAEAKKIIGITKNNDIWKLDLGHATEAVOALOERLQ  
IEGFTYRVSYPNRKSKNAALASOEPPGRILDEEGHGCORTSRELANSLQVITG  
IGKHSRQASLPLAVKTFEDNRYRFBETPRGVITVYKPRHS"  
complement(12417..13424)  
/note="contains similarity to unknown protein  
gb|AAF27018.1  
gene\_id:MOM1.21"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="BAA97248.1"  
/db\_xref="GI:8809707"  
/translation="WATDSQPNOKLTLPLKTRIALTVISTWTDNAORPDGTINERFLK  
LEDFRAPNPKEVNIYSTDPVQSDRLMFLTYTPVSGDKIPVVFEGGGAFIS  
PNAYPVNCRPARKLPAYIVSNVRLAPRPAQYDGFALKIENHGSILPA  
NADLSRCFFAGDSAGNIAHVAVIRICEPSSFTAVLIGLISIOFEGGEETAE  
KQLVGAPLVSPTRIDWCKAKGLNRDHAUVANGPNAVDISGLDYPETMVVAGFDL  
KDWRSYEWMLCGKATLLIYPMWFAFTIFPLPAGOLINRIKDFADERAASIS  
A"  
join(14922..14981,15072..15167,15563..15666,15755..15891,  
16096..16178)  
/note="gene\_id:MOM1.20"  
/codon\_start=1  
/evidence=not experimental  
/product="50S ribosomal protein L24"  
/protein\_id="BAA97247.1"

/db\_xref="GI:8809706"  
/translation="MGWKAELIRHMKILRGDVMILRGDKGEFTIKRVIRSONR  
VTEGKULAKH1KGGPHEGGITPVEALHNSNVQVDPVGRCKGVKYLEGDK  
VVARGTGTSSTIPEPELTKIRATPRPTAPKQTPMEFWEQYDAKTGKGMPL"  
join(16823..16952,17143..17320,17412..17493,17675..17851,  
17943..18110,18199..18390)  
/note="gene\_id:MOM1.19"  
/codon\_start=1  
/evidence=not experimental  
/product="26S proteasome, non-ATPase regulatory subunit"  
/protein\_id="BAA97246.1"  
/db\_xref="GI:8809705"  
/translation="MERLQIRFAGGGLGHASPDFTLTSQVYISSIALKMLKRG  
RACVPMWGLMGEFVDETVAVDVVFAMPQSGVSVSAVDHVFQTMMLMKQTG  
RPMVAVGHSHFGPGCNSVDINTQSFALNQAVAVVDPLOSXGVKVIDAFR  
SINPOTIMGQREORTNSLGHNRKSIQALHGLRHYSLIAINRKNELEKMLN  
LHKRWTDGLTLRFTHTSKTNEQTOEMLSLAAKKNKAVQEBDELSPKLA1VNVGR  
QDAKRLIEHVSIMSSNIVQTLGTLDTVP"  
complement(join(18965..19043,19240..19238,19419..19536,  
19666..19739,19908..19961,20387..20530))  
/note="contains similarity to unknown protein  
emb|CAA2897.1  
gene\_id:MOM1.18"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="BAA97245.1"  
/db\_xref="GI:8809704"  
/translation="MDKNOAIEKMKLVGMEVEDEQAADSESSISFMEDLNRCAL  
TTKORFYGRALCSAGITCTLLSMVFPNVKRGITPTGLNALGSTAFILIQPOV  
TMLDPAKRYATLALYLASITIALFCALYVANKLLTLLAILEFTGLIWSLSY1PFR  
TMSKIFMTCEDEIE"  
join(22672..23634,23708..23999,24111..24355,24458..24624,  
24715..24925)  
/note="emb|CAB62356.1  
gene\_id:MOM1.17  
similar to unknown protein"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="BAA97244.1"  
/db\_xref="GI:8809703"  
/translation="MSSRAGPMSKKNVQGGYRVEBOLVOGLAGTLASODDGERM  
EVSIRKKNKKNKNTSGKTWYNSGNPNRPNAGGQGGGVSGVGNVSGNNGNG  
IQANISGRGALSKRTDNNVAFPPVYRPPLEGGMNOAGGSAOHTAVOEPVEDD  
VDNASBEENDSLDSDDDLASDYSDVSQKSHGRKONKWKPKKFGSISLSTEO  
INEPQOWHCPACQNGPGLADWYIIVPVPVITNTRLDKODNKKMLKMGQELIEF  
ASVIECGE1YQOMKGLGEDEKDYBIIVPVPVITNTRLDKODNKKMLKMGQELIEF  
DKYELRABHSGYQGHKMGSVLMFESATGYLEAERLHRELAEMGLDR1AMQGRM  
PSCYRQIYGLATRKODIDINOSHOGKTLKFLKSKJOEYVAVELKQISDNQOLY  
FKNLSKONKAKVLESE1TMSKLRTRTAEADRIVORTMOHOREEMDADRFF  
MDSIKQIHERDPAKEENBEM1QOQERAKVVGQOOONINPSNDNCCRRAREVSFIEP  
DEKEMEERFEREMLIKQEKEMEDMKRHHLE1FDEKEFDELBQMLYHGLAHND  
D"  
join(25614..25838,26061..26207,26324..26438,26516..26610,  
26710..26813,27522..27765,27943..28036,28115..28305,  
28385..28604,28688..28855,28944..29092)  
/note="gb|AAC97420.1  
gene\_id:MOM1.16  
strong similarity to unknown protein"  
/codon\_start=1  
/evidence=not experimental  
/protein\_id="BAA97243.1"  
/db\_xref="GI:8809702"  
/translation="WAPPAGQTAVAEAAGDAPPOOORRGSTISGIVRIAVWFY  
ASKFSPKQKMPDPTAPQLMTNLPHKESLDWVYLSQEKENDFGNDALYHETN  
IPAAWTPESIRKTLIYVPSPTLONNSLYAH1FFASGEPIDTDEYOPUNFSR  
THAVVYEPKOKNKKSLGSPKDSDESEPEVEKVGDKSPKEVEVEM1SLKPN  
VTINLVDPDTPADLILVEPTGNVYPTIYFNEFLIRDKFIPVNEVSELPLNTEIS  
ISMKQQLQOQDQSFQWORSYGMWLDSESDSLKVPLEGNVYLLGIMFVSMHSVF  
DLAKNDIQIFWNNKNSWEGSASKAVYALIQPITFLYLLNDNSMTLASGVGC  
IEFWKIGKAMRIEVRSGMIPLRPHDRASVANSKTRKEDDAIKFLSVILLIYIGL  
SIYSIAVERHKSMTLSLSTSCVMYGFIMWCQLFLNYLKLKVAHLPMQMTYK  
LNTIIDLPAFYIKMPLIHLRLSVFRDDVFLIYLVQWVYVDPKTRVNEFGGDEET  
ARKCLITEKEEDKKTN"

CDS		complement(join(29439..29576,29661..29885,29985..30268, 30341..30493,30579..30722,30956..31484))		/note="gene_id:MOM1.15"	
Query Match	100.0%;	Score 3275;	DB 8;	Length 81365;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 3275;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	GACAAACAAACAAATTAAGCAAGTCATGTCGTAGCAATAATTAATAGTGGGAAACAA	60		
DB	21977	GACAAACAAACAAATTAAGCAAGTCATGTCGTAGCAATAATTAATAGTGGGAAACAA	22036		
QY	61	TTAAGTTAAGCGAAAAAGAAAAAGGTACAAAAATGAAAAACAATCAACTGAAT	120		
DB	22037	TTAAGTTAAGCGAAAAAGAAAAAGGTACAAAAATGAAAAACAATCAACTGAAT	22096		
QY	121	GAAAAATTTGGAGTCGCAAGATCGGAAAAAGAGCCGTTTAAAGCTTAATAGCTTC	180		
DB	22097	GAAAAATTTGGAGTCGCAAGATCGGAAAAAGAGCCGTTTAAAGCTTAATAGCTTC	22156		
QY	181	ATTGTCCTCTCTGTCAGGTTTATTTCTTCCCGAGTCCTGACTCACTACTCAC	240		
DB	22157	ATTGTCCTCTCTGTCAGGTTTATTTCTTCCCGAGTCCTGACTCACTACTCAC	22216		
QY	241	TCGCCGCGCTTAACTTAAGTTCCTCCGTCCTTAATCTCTGAAGTTTCTGCCCTAGAG	300		
DB	22217	TCGCCGCGCTTAACTTAAGTTCCTCCGTCCTTAATCTCTGAAGTTTCTGCCCTAGAG	22276		
QY	301	CTCCGATGCGCTCACCGCATGCAATCTGTGCTCGAATTTCTTTTCTTCCGTGAAAA	360		
DB	22277	CTCCGATGCGCTCACCGCATGCAATCTGTGCTCGAATTTCTTTTCTTCCGTGAAAA	22336		
QY	361	ATTGCCCTAATGTCGATTTGGAAGTTTGTGCTAAGGGTTACTTTTCCCTATA	420		
DB	22337	ATTGCCCTAATGTCGATTTGGAAGTTTGTGCTAAGGGTTACTTTTCCCTATA	22396		
QY	421	TTTTATAGTCTTAGGTAAAGATACCTGCTCTTACTGTTTGTCTCAATTTGTGTGCT	480		
DB	22397	TTTTATAGTCTTAGGTAAAGATACCTGCTCTTACTGTTTGTCTCAATTTGTGTGCT	22456		
QY	481	TTTCAACGTTTATGCTGATCGGATCGGAGTATTGACTGGAAAAATCCTTCGTTTGTGTT	540		
DB	22457	TTTCAACGTTTATGCTGATCGGATCGGAGTATTGACTGGAAAAATCCTTCGTTTGTGTT	22516		
QY	541	TTGTTTCATATAAATGGAATGATCTACCTTTGTGCTTTGATGTTTGTGAGCCT	600		
DB	22517	TTGTTTCATATAAATGGAATGATCTACCTTTGTGCTTTGATGTTTGTGAGCCT	22576		
QY	601	ATGCGTGTGTGCTGTATACTTCACTTCACGTTCATGTGTGATTTGAGATTGTGATG	660		
DB	22577	ATGCGTGTGTGCTGTATACTTCACTTCACGTTCATGTGTGATTTGAGATTGTGATG	22636		
QY	661	ACTGTGGGTTTCTTTGTGCTATAGGTTGTAAAAATGAGTTCTAGGGCTGCTCATGT	720		
DB	22637	ACTGTGGGTTTCTTTGTGCTATAGGTTGTAAAAATGAGTTCTAGGGCTGCTCATGT	22696		
QY	721	CTAAGGAAAAAGACGTTCAAGGTTTATAGGCTTAGGTTGAACGTTGTTCAAGGTT	780		
DB	22697	CTAAGGAAAAAGACGTTCAAGGTTTATAGGCTTAGGTTGAACGTTGTTCAAGGTT	22756		
QY	781	TGGCAGGAGCAGACTGGCTTCTTCAACAAGTATGAGAGAGAGTGGAGTCAATTCCA	840		
DB	22757	TGGCAGGAGCAGACTGGCTTCTTCAACAAGTATGAGAGAGAGTGGAGTCAATTCCA	22816		
QY	841	AGAAGAACAAGAACCAACGAGAAACATTTGGAACAACTGGGTTTCTCAGAATTCCA	900		
DB	22817	AGAAGAACAAGAACCAACGAGAAACATTTGGAACAACTGGGTTTCTCAGAATTCCA	22876		
QY	901	ATCCTCTTAGAGCTTGGGGTGTGACAGACAGGAGAGGTAGCAAGTATCTGGGAGAG	960		
DB	22877	ATCCTCTTAGAGCTTGGGGTGTGACAGACAGGAGAGGTAGCAAGTATCTGGGAGAG	22936		
QY	961	GAAACAATGTATTCGGAGAGGTAAACGGCAATGTGTGGGGCAATTCAAGCTAAATATCTG	1020		

DB	22937	GAAACAATGTATTCGGAGAGGTAAACGGCAATGTGTGGGGCAATTCAAGCTAAATATCTG	22996		
QY	1021	GTCGGGAGCAGCGTTGACACAAAAGTATGATTAACAATTTGTGGACACCCCACTGTAT	1080		
DB	22997	GTCGGGAGCAGCGTTGACACAAAAGTATGATTAACAATTTGTGGACACCCCACTGTAT	23056		
QY	1081	CTCGCCCTCTTTGGAAGAGATGGAATTGGCAGGCAAGAGAGGTTCTGTCAACACA	1140		
DB	23057	CTCGCCCTCTTTGGAAGAGATGGAATTGGCAGGCAAGAGAGGTTCTGTCAACACA	23116		
QY	1141	CAGCTGTGAGAGATTCTCTGACGTGAGAGATGATGTGATTAATGCTTGTGAGAAAGAC	1200		
DB	23117	CAGCTGTGAGAGATTCTCTGACGTGAGAGATGATGTGATTAATGCTTGTGAGAAAGAC	23176		
QY	1201	ATGATTCGATGCTTTGATGATGATCTGATGACGACCTTCGAAGTATGATGATCGG	1260		
DB	23177	ATGATTCGATGCTTTGATGATGATCTGATGACGACCTTCGAAGTATGATGATCGG	23236		
QY	1261	ATGTAGTCAAAAGAGCCATGATCAAGAAAGACAAATAGTGTCAAAAAGTCTTGG	1320		
DB	23237	ATGTAGTCAAAAGAGCCATGATCAAGAAAGACAAATAGTGTCAAAAAGTCTTGG	23296		
QY	1321	GCAAGTTGATAGCTTGTGATCGAGCAGATTAATGAACACAGAGCAGTGGCAATTGTC	1380		
DB	23297	GCAAGTTGATAGCTTGTGATCGAGCAGATTAATGAATGAACACAGAGCAGTGGCAATTGTC	23356		
QY	1381	CAGCTTGACAAACGACCTGTCATGATGATTTGGTATTAACCTGCACCTCTACTAGCTC	1440		
DB	23357	CAGCTTGACAAACGACCTGTCATGATGATTTGGTATTAACCTGCACCTCTACTAGCTC	23416		
QY	1441	ATGCGAGCACAAGAGAGCTAGGCGAGTTAAGCTCATAGAAATGGCTGAAGTTTAG	1500		
DB	23417	ATGCGAGCACAAGAGAGCTAGGCGAGTTAAGCTCATAGAAATGGCTGAAGTTTAG	23476		
QY	1501	AAAAGATCTACAGATGAGAGCGCATCTGTCAATTCCTTGTGTGATTTAATGGCAGT	1560		
DB	23477	AAAAGATCTACAGATGAGAGCGCATCTGTCAATTCCTTGTGTGATTTAATGGCAGT	23536		
QY	1561	GGAAGGTTTGGGTGAGATGAAAAAGATTATGAATTTGTCTGGCTCCAAATGGTATCA	1620		
DB	23537	GGAAGGTTTGGGTGAGATGAAAAAGATTATGAATTTGTCTGGCTCCAAATGGTATCA	23596		
QY	1621	TCATGAATCTAGACCTGATTAAGACGATPAAGGTGAATTTCTCTGCTTTTAC	1680		
DB	23597	TCATGAATCTAGACCTGATTAAGACGATPAAGGTGAATTTCTCTGCTTTTAC	23656		
QY	1681	TTCTTAAATTTTCTTTCATTTCTAGATCTGATCTAGAAATGTATCATGTAGTGGCTCGGC	1740		
DB	23657	TTCTTAAATTTTCTTTCATTTCTAGATCTGATCTAGAAATGTATCATGTAGTGGCTCGGC	23716		
QY	1741	ATGGGACACCAAGAGCTGTGGAATTACTTCACAAGATAGAGGCTTTAGACAGCCAT	1800		
DB	23717	ATGGGACACCAAGAGCTGTGGAATTACTTCACAAGATAGAGGCTTTAGACAGCCAT	23776		
QY	1801	TCCTAATGTCACAGGGCCATGTTGGAGTGAATGATGTTTCTGATGTTTGAAGAGCAGTGCAC	1860		
DB	23777	TCCTAATGTCACAGGGCCATGTTGGAGTGAATGATGTTTCTGATGTTTGAAGAGCAGTGCAC	23836		
QY	1861	GGCTAATTTGGAGCCGAACGCTCCACCGGAGTTAGCTGAGATGGGGTTAGATGAATT	1920		
DB	23837	GGCTAATTTGGAGCCGAACGCTCCACCGGAGTTAGCTGAGATGGGGTTAGATGAATT	23896		
QY	1921	GCTGTGGGTGACAAACGCAATGATTTTCTGAGAGTGTTCGCAACTGATATGGCTTCTT	1980		
DB	23897	GCTGTGGGTGACAAACGCAATGATTTTCTGAGAGTGTTCGCAACTGATATGGCTTCTT	23956		
QY	1981	GCAAGAAACCAAGCTGACATTAATCAACCTCTCAAGTTCCTCTCCCAAGAA	2040		
DB	23957	GCAAGAAACCAAGCTGACATTAATCAACCTCTCAAGTTCCTCTCCCAAGAA	24016		
QY	2041	AATTTGATATATGCTTTAGTTTGTATGGAATTTAAAGTTTGTGTGCTCGGTATA	2100		

```

Db      24017 AATTGATATATGCTTTAGTTTGTTCATGGAATTTAAAGTTTGTGTCCTGTTAA 24076
Qy      2101 TGCATCTGTATGTAATATATCATGATTCATAGGCAAAACAGGCTGAATTCGAGTTG 2160
Db      24077 TGCATCTGTATGTAATATATCATGATTCATAGGCAAAACAGGCTGAATTCGAGTTG 24136
Qy      2161 AATCATACCAAGAGATGTTGTAAAGAGCTGAGGAGATCTTGAGGACATACAGAG 2220
Db      24137 AATCATACCAAGAGATGTTGTAAAGAGCTGAGGAGATCTTGAGGACATACAGAG 24196
Qy      2221 CTGAATCTATTTAAGAACAGCTCTCAAAACAGAACAGGCGCAAGGTCCTGAGGA 2280
Db      24197 CTGAATCTATTTAAGAACAGCTCTCAAAACAGAACAGGCGCAAGGTCCTGAGGA 24256
Qy      2281 TCTCTGAAATTTATGAGCAGAGAGCTGCTAGAACCTGACAGAGATTAATGAGTCGTAGA 2340
Db      24257 TCTCTGAAATTTATGAGCAGAGAGCTGCTAGAACCTGACAGAGATTAATGAGTCGTAGA 24316
Qy      2341 CAGAGAACTAAGATGACAGATGAAACAGAACAGGAGATGATTTTCTTAGAAAT 2400
Db      24317 CAGAGAACTAAGATGACAGATGAAACAGAACAGGAGATGATTTTCTTAGAAAT 24376
Qy      2401 CAGAACTTGACATTTTGTATTAATCACTGATTCACATTTTGTATTAATGTCACAA 2460
Db      24377 CAGAACTTGACATTTTGTATTAATCACTGATTCACATTTTGTATTAATGTCACAA 24436
Qy      2461 AAAACCTGTGCTGTTGAAGATGATGACACAGAGGTTTTCATGATTCATCAAA 2520
Db      24437 AAAACCTGTGCTGTTGAAGATGATGACACAGAGGTTTTCATGATTCATCAAA 24496
Qy      2521 CAGATCCATGAAGAGAGAGCGCAAGAGAGAGATTTTGAGATGTTTGACAGCAGAGAA 2580
Db      24497 CAGATCCATGAAGAGAGAGCGCAAGAGAGAGATTTTGAGATGTTTGACAGCAGAGAA 24556
Qy      2581 CGGCGCAAGGTTGTGGCCAGCAGCAGCAGCAGACATTAATCCCTCTAGCATAGCAGTTGC 2640
Db      24557 CGGCGCAAGGTTGTGGCCAGCAGCAGCAGCAGACATTAATCCCTCTAGCATAGCAGTTGC 24616
Qy      2641 CGAAGAGGTATATGTAATTAATCACTAATCACTAATCCCTCTGCGCTTTTGTTCCTA 2700
Db      24617 CGAAGAGGTATATGTAATTAATCACTAATCACTAATCCCTCTGCGCTTTTGTTCCTA 24676
Qy      2701 AGAGTACTGAATTAATCCGCTTTGATCTTTGCGAGCTGAGAGAGTCAAGCTTC 2760
Db      24677 AGAGTACTGAATTAATCCGCTTTGATCTTTGCGAGCTGAGAGAGTCAAGCTTC 24736
Qy      2761 ATCGAGTTTCAAGAGAAAGAGATGAGAGAGTGTGTGAGAGAGAGAGATGCTGATPAAA 2820
Db      24737 ATCGAGTTTCAAGAGAAAGAGATGAGAGAGTGTGTGAGAGAGAGAGATGCTGATPAAA 24796
Qy      2821 GATCAAGAGAGAGATGAGAGACATGAAGAGAGGCTCAACAGAGATTAATTTGATCTG 2880
Db      24797 GATCAAGAGAGAGATGAGAGACATGAAGAGAGGCTCAACAGAGATTAATTTGATCTG 24856
Qy      2881 GAGAGAGATTTGATGAGAGCTTTGAGACAGCTCATGTAACAAGATGCTTCACATGAA 2940
Db      24857 GAGAGAGATTTGATGAGAGCTTTGAGACAGCTCATGTAACAAGATGCTTCACATGAA 24916
Qy      2941 GATGATTTGAGACAAAGTCTGTGACACAGACAAAGTCTTAAGTCTTTGCTTTG 3000
Db      24917 GATGATTTGAGACAAAGTCTGTGACACAGACAAAGTCTTAAGTCTTTGCTTTG 24976
Qy      3001 GTATGTCGAAAGTGAAGATCTGAGAGCTCCATTTAAATCTAGAGCAAACTTAAGGA 3060
Db      24977 GTATGTCGAAAGTGAAGATCTGAGAGCTCCATTTAAATCTAGAGCAAACTTAAGGA 25036
Qy      3061 GATTATATGATTAATTCCTCCAAATTTTATGAGACGATCTAAGAGACATTAAGTCTT 3120
Db      25037 GATTATATGATTAATTCCTCCAAATTTTATGAGACGATCTAAGAGACATTAAGTCTT 25096
Qy      3121 GTGACTAAACCAAGTTTCTTATGATTTTGTTCCTTTTGTTCCTTTTGTTCCTTTTGTTC 3180
Db      25097 GTGACTAAACCAAGTTTCTTATGATTTTGTTCCTTTTGTTCCTTTTGTTCCTTTTGTTC 25156

```

```

Qy      3181 TTAGACATATTACCAACGTCAGAGTGAATCAAGAAATGGCAATCAAAATCATGTTTT 3240
Db      25157 TTAGACATATTACCAACGTCAGAGTGAATCAAGAAATGGCAATCAAAATCATGTTTT 25216
Qy      3241 AGAATTTATATCTACAAATATATGATGATCAAAAT 3275
Db      25217 AGAATTTATATCTACAAATATATGATGATCAAAAT 25251

RESULT 3
AF239719      2254 bp      DNA      linear      PLN 02-JUN-2000
LOCUS      Arabidopsis thaliana SGS3 gene, complete cds.
DEFINITION      Arabidopsis thaliana SGS3 gene, complete cds.
ACCESSION      AF239719
VERSION      AF239719.1 GI:8164029
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 2254)
Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, P., Godon, C.,
Morel, J.B., Jonette, D., Lacombe, A.M., Nikic, S., Picault, N.,
Remoue, K., Santal, M., Vo, T.A. and Vaucheret, H.
Arabidopsis SGS2 and SGS3 genes are required for
posttranscriptional gene silencing and natural virus resistance
Cell 101 (5), 533-542 (2000)
JOURNAL
MEDLINE
20306668
PUBMED
10850495
2 (bases 1 to 2254)
Beclin, C., Mourrain, P., Vaucheret, H. and Elmayan, T.
Direct Submission
JOURNAL
TITLES
Submitted (28-FEB-2000) Biologie Cellulaire, INRA, Route de
Saint-Cyr, Versailles 78026, France
FEATURES
source
1..2254
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="V"
join(<1..963,1037..1328,1440..1684,1787..1953,2044..>2254)
/product="SGS3"
join(1..963,1037..1328,1440..1684,1787..1953,2044..2254)
/codon_start=1
/product="SGS3"
/protein_id="AA173960.1"
/translation="MSRAAGPMSEKKNVVOGYRPEBQIVQGLAGTTLASQDQGGN
EVISSKNNKPKPNTSGKTNVQNSNPVPMGQOQOGRSNVSGKNNVSGKNGRG
IOANTSGRGLSRKYDNFVAPVPSPLEGEGWAGRGSAQHTVVOEPVEDD
VNASEENDSDALDDSDLDSDYDSVQKSGSRKQNKWFKKPFGLSLSIQ
INEROMHCPCQNCQNSALIDPMYNIPLAARTRGABRVLRHLAVLEKDLQMRG
ASVLCGEIYGQMKLGEDQYELVWPMTVIMTRLDKQNDKMLGMQOELLEF
DYREALRARSYGPQGRHGMVLMESATGILELRHLRLAEKGLDRIMQGRSM
FSGYRQLYGLFATQODLIFVQHGQRTLRKFLKSYQENVVKELRQISDNOQNT
FNKLSKONKAKLVLESLEIWEKLRRTAEDNRIVRQRTQOHEONRENDADHDF
MSTQKHERRRRAKNEPMLQQORAVVQGOONINPSSDDCRKAERYSSIEF
QEKEMBEFVEREMLIKQEKMEMKRRHHEIFLDEKFEPALEQLMYKGLHNE
D"

ORIGIN
Query Match      68.8%; Score 2254; DB 8; Length 2254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      696 ATGAGTTCTAGGGCTGTCATGCTTAAGAGAAAGACGTTCAAGGCTGTTATAGCCT 755
Db      1 ATGAGTTCTAGGGCTGTCATGCTTAAGAGAAAGACGTTCAAGGCTGTTATAGCCT 60
Qy      756 GAGGTGAACAGTTGTTCAAGTTTGCGAGGAGAGAGAGCTGGCTTCTCACAAATGAT 815

```



```

Db      ||||| 61 GAGGTGAACAGTGGTTCAAGGTTGGCAGGAGCAGACTGGCTTCTTCAAGAATGAT 120
Qy      ||||| 816 GAGAGAGAGTGGAGGTCATTTTCCAAAGAAACAACAACAACCAAGAAACATTTCTGCA 875
Db      ||||| 121 GAGAGAGAGTGGAGGTCATTTTCCAAAGAAACAACAACAACCAAGAAACATTTCTGCA 180
Qy      ||||| 876 AAAAATTGGGTTCTCAGAAATTCCTCTAGAGCTTGGGTGGTCCAGCAAGG 935
Db      ||||| 181 AAAAATTGGGTTCTCAGAAATTCCTCTAGAGCTTGGGTGGTCCAGCAAGG 240
Qy      ||||| 936 AGAGTAGCAACGTAATCTGGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATGGT 995
Db      ||||| 241 AGAGTAGCAACGTAATCTGGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATGGT 300
Qy      ||||| 996 CGGGGATTCAGACTAACATTTCTGGTCCGGGACGAGCTTGGACAGAAATGTATTAAC 1055
Db      ||||| 301 CGGGGATTCAGACTAACATTTCTGGTCCGGGACGAGCTTGGACAGAAATGTATTAAC 360
Qy      ||||| 1056 AACTTGTGGCAACCCCACTGTATCTGGCTCTCTTGGAGAGAGATGGAATTTGGCAG 1115
Db      ||||| 361 AACTTGTGGCAACCCCACTGTATCTGGCTCTCTTGGAGAGAGATGGAATTTGGCAG 420
Qy      ||||| 1116 GCAAGAGAGAGTTCCTCTCAGCAACAACCTGTCCAGAGATTTCTGAAGTGAATGAT 1175
Db      ||||| 421 GCAAGAGAGAGTTCCTCTCAGCAACAACCTGTCCAGAGATTTCTGAAGTGAATGAT 480
Qy      ||||| 1176 GTGGATTAATCTTCTGAGAGAGAAATGATCCGATTCCTTGGATGATTTGTATGACAC 1235
Db      ||||| 481 GTGGATTAATCTTCTGAGAGAGAAATGATCCGATTCCTTGGATGATTTGTATGACAC 540
Qy      ||||| 1236 CTTGCAAGTGAATGATTAATGATCTCGAGATGTGAGTCAAAAGAGCCATGGATCAGAAAGCAG 1295
Db      ||||| 541 CTTGCAAGTGAATGATTAATGATCTCGAGATGTGAGTCAAAAGAGCCATGGATCAGAAAGCAG 600
Qy      ||||| 1296 AATTAAGTGTTCAAAAGATCTTTTGGCAGCTTGGATGATCTTGTGATTCAGACAGATAAAT 1355
Db      ||||| 601 AATTAAGTGTTCAAAAGATCTTTTGGCAGCTTGGATGATCTTGTGATTCAGACAGATAAAT 660
Qy      ||||| 1356 GAACCAAGAGGAGTGGCATTGTCCAAGCTTGTCAAGACGAGCTGTGCATTCGATTTGG 1415
Db      ||||| 661 GAACCAAGAGGAGTGGCATTGTCCAAGCTTGTCAAGACGAGCTGTGCATTCGATTTGG 720
Qy      ||||| 1416 TATTAACCTGACCTCTACTAGCTCATGAGAGCAAAAGAGACTGGGAGTAAAGCTC 1475
Db      ||||| 721 TATTAACCTGACCTCTACTAGCTCATGAGAGCAAAAGAGACTGGGAGTAAAGCTC 780
Qy      ||||| 1476 CATAGAGAAATGGCTGAAGTTTATAGAAAAGATCTACAGATGAGAGCGCATCTGCATT 1535
Db      ||||| 781 CATAGAGAAATGGCTGAAGTTTATAGAAAAGATCTACAGATGAGAGCGCATCTGCATT 840
Qy      ||||| 1536 CTTTGTGTGAGATTTATGTGGCAGTGAAGGTTTGGGTGAGATGAAAAGATTGAA 1595
Db      ||||| 841 CTTTGTGTGAGATTTATGTGGCAGTGAAGGTTTGGGTGAGATGAAAAGATTGAA 900
Qy      ||||| 1596 AATTGTGGCTCAATGGTCATCATGATTAAGTAACTGATGATTAAGAGCAATGAGAT 1655
Db      ||||| 901 AATTGTGGCTCAATGGTCATCATGATTAAGTAACTGATGATTAAGAGCAATGAGAT 960
Qy      ||||| 1656 AAGGTGAATTTCTTGTCTTTTACTTCTTAAATTTTCTCTTGCATTTCTACTGATCTTA 1715
Db      ||||| 961 AAGGTGAATTTCTTGTCTTTTACTTCTTAAATTTTCTCTTGCATTTCTACTGATCTTA 1020
Qy      ||||| 1716 GAATTTTAACTTTATAGTGGCTCGGCATGGGCAACAAGAGCTCTGGAATCTTCCACAA 1775
Db      ||||| 1021 GAATTTTAACTTTATAGTGGCTCGGCATGGGCAACAAGAGCTCTGGAATCTTCCACAA 1080
Qy      ||||| 1776 GTATGAGGCTTTAGAGCAAGCAATTCATATGATTCACAGGGCATGTGGAGTGAATGT 1835
Db      ||||| 1081 GTATGAGGCTTTAGAGCAAGCAATTCATATGATTCACAGGGCATGTGGAGTGAATGT 1140
Qy      ||||| 1836 TCTGATGTTTGAAGAGATGCCATGCTATTTTGAAGGCCGAACGCTCCACCGGAGTT 1895

```

---

```

Db      ||||| 1141 TCTGATGTTTGAAGAGATGCCATGCTATTTTGAAGGCCGAACGCTCCACCGGAGTT 1200
Qy      ||||| 1896 AGCTGATGGGTTTATGATGATTTGCTGGGGTCAGAACCGCAGATGTTTCTGAGG 1955
Db      ||||| 1201 AGCTGATGGGTTTATGATGATTTGCTGGGGTCAGAACCGCAGATGTTTCTGAGG 1260
Qy      ||||| 1956 TGTTCGCAACTGATATGAGCTTCTTCCACGAAGCAAGATCTGACATATTCATCAACA 2015
Db      ||||| 1261 TGTTCGCAACTGATATGAGCTTCTTCCACGAAGCAAGATCTGACATATTCATCAACA 1320
Qy      ||||| 2016 CTCTCAAGTTCTCTCCCAAGAAATTTGATATATGCTTTTACTTTTGTCTGGAAT 2075
Db      ||||| 1321 CTCTCAAGTTCTCTCCCAAGAAATTTGATATATGCTTTTACTTTTGTCTGGAAT 1380
Qy      ||||| 2076 TTAAGTTTGTGTCGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 2135
Db      ||||| 1381 TTAAGTTTGTGTCGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
Qy      ||||| 2136 CAAGAGAGTGAATTCAGATTTGAATCATACCAAGAGATGTTTGAAGAGCTGAG 2195
Db      ||||| 1441 CAAGAGAGTGAATTCAGATTTGAATCATACCAAGAGATGTTTGAAGAGCTGAG 1500
Qy      ||||| 2196 GCAGATCTTGAAGCAATCAGAGCTGAACCTTTAAGAACAGCTCTCAAAACAGAA 2255
Db      ||||| 1501 GCAGATCTTGAAGCAATCAGAGCTGAACCTTTAAGAACAGCTCTCAAAACAGAA 1560
Qy      ||||| 2256 CAAGAGAGTGAATTCAGATTTGAATCATACCAAGAGATGTTTGAAGAGCTGAG 2315
Db      ||||| 1561 CAAGAGAGTGAATTCAGATTTGAATCATACCAAGAGATGTTTGAAGAGCTGAG 1620
Qy      ||||| 2316 TCGAGAGATTAATCGGATCTGAGACAGAGAACTAAGATGACAGATGAACAGAGGA 2375
Db      ||||| 1621 TCGAGAGATTAATCGGATCTGAGACAGAGAACTAAGATGACAGATGAACAGAGGA 1680
Qy      ||||| 2376 AAGATATATTTTCTTGAAGAAATCAAACTTGCATTTTGTATTAATCTACTGATCA 2435
Db      ||||| 1681 AAGATATATTTTCTTGAAGAAATCAAACTTGCATTTTGTATTAATCTACTGATCA 1740
Qy      ||||| 2436 CATTTTGAATTAATGTCACAAAGAAACCTGTGTGTGTTGAAGATGACACAGA 2495
Db      ||||| 1741 CATTTTGAATTAATGTCACAAAGAAACCTGTGTGTGTTGAAGATGACACAGA 1800
Qy      ||||| 2496 CAGTTTTCATGATTCATCAACAGATTCATGAAGAAAGAGAGCAAGAGAGAGAA 2555
Db      ||||| 1801 CAGTTTTCATGATTCATCAACAGATTCATGAAGAAAGAGAGCAAGAGAGAGAA 1860
Qy      ||||| 2556 TTTGAGATGTTGCAGACAGAGAGCTGCAAGTGTGTGCGCAGACAGACAGAAAT 2615
Db      ||||| 1861 TTTGAGATGTTGCAGACAGAGAGCTGCAAGTGTGTGCGCAGACAGACAGAAAT 1920
Qy      ||||| 2616 TAACTCTTACGATGAGATTTGCGGAAAGAGTATATGTAATCAATTAATCCCT 2675
Db      ||||| 1921 TAACTCTTACGATGAGATTTGCGGAAAGAGTATATGTAATCAATTAATCCCT 1980
Qy      ||||| 2676 CTGGCGTTTGTTTTCAAACTTAAGATTAATTAATCCGGTTTGAATCTTTG 2735
Db      ||||| 1981 CTGGCGTTTGTTTTCAAACTTAAGATTAATTAATCCGGTTTGAATCTTTG 2040
Qy      ||||| 2736 CAGAGCTGAGAGATGTCAGCTTCAATGAGTTTCAAGAGAAAGAGATGAGAGATTTGT 2795
Db      ||||| 2041 CAGAGCTGAGAGATGTCAGCTTCAATGAGTTTCAAGAGAAAGAGATGAGAGATTTGT 2100
Qy      ||||| 2796 GGAAGAGAGAGATGTCATTAAGATTAAGAGAAAGAGATGAGAGATGAGAGAG 2855
Db      ||||| 2101 GGAAGAGAGAGATGTCATTAAGATTAAGAGAAAGAGATGAGAGATGAGAGAG 2160
Qy      ||||| 2856 GCATCAAGAGATATTTGATCTGGAAGAAATTTTGAAGGCTTTGGAACAGCTCAT 2915
Db      ||||| 2161 GCATCAAGAGATATTTGATCTGGAAGAAATTTTGAAGGCTTTGGAACAGCTCAT 2220
Qy      ||||| 2916 GTTCAAGATGAGCTTCAATGAAGATGATGA 2949
Db      ||||| 2221 GTTCAAGATGAGCTTCAATGAAGATGATGA 2254

```



RESULT 4  
BT002944  
LOCUS  
DEFINITION  
Arabidopsis thaliana clone RAF14-93-K05 (R20243) unknown protein  
(AT5G23570) mRNA, complete cds.

ACCESSION  
BT002944  
BT002944.1 GI:27754622

VERSION  
FLI CDNA.

KEYWORDS  
Arabidopsis thaliana (thale cress)

SOURCE  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eustosida II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 2162)

REFERENCE  
AUTHORS  
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,  
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,  
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,  
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,  
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,  
Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,  
Ecker, J.R. and Theologis, A.  
Arabidopsis Full Length cDNA Clones  
Unpublished  
2 (bases 1 to 2162)

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,  
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,  
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,  
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,  
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,  
Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,  
Ecker, J.R. and Theologis, A.  
Direct Submission  
Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAFL cDNAs (RAFL cDNA: 'Ishida, J.,  
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y. and Shinzaki, K.

COMMENT  
The Salk, Stanford, PEGC (SSP) Consortium members carried out the  
sequencing and annotation of the RAFL cDNAs: Yamada, K., Chan, M.M.,  
Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Onodera, C.S.,  
Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yu, G.,  
Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M.,  
Palm, C.J., Shim, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W.,  
Ecker, J.R. and Theologis, A.  
Yamada, K. (SSP/PEGC) and Seki, M. (RIKEN GSC) contributed equally to  
this work. Shinzaki, K. (RIKEN GSC) and Theologis, A. (SSP/PEGC)  
contributed equally to this work as PIs.

FEATURES  
Source  
Annotation based on July 2002 version of the Arabidopsis genome  
submitted to Genbank.  
Location/Qualifiers  
1..2162  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="RAF14-93-K05 (R20243)"  
/ecotype="Columbia"  
/note="This clone is in a modified pBluescript vector2  
(lambda PS) as a BamHI/XhoI insert."  
1..2162  
/gene="At5g23570"  
1..91  
/gene="At5g23570"  
92..1969  
/gene="At5g23570"  
/codon\_start=1

gene  
5' UTR  
CDS

ORIGIN  
Query Match 29.7%; Score 972; DB 8; Length 2162;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
misc\_difference 2146  
/gene="At5g23570"  
/note="not present in genomic sequence"

3' UTR  
misc\_difference 2146  
/gene="At5g23570"

1970..2162  
/gene="At5g23570"

/evidence=experimental  
/product="unknown protein"  
/protein\_id="AA022757.1"  
/db\_xref="GI:27754623"  
/translation="MSRAPPMSEKRVVGGYRPEVQLVQGLATPTLASSODNGEM  
EVLSKRNKRGNTSGKTWVSQNSNPRAVGGQGGSVSGRNVSGRKNNGG  
IQANISGRGRALSRKYDNNFVAPPPVSRPLPEGGMNQAGGSAKFFAVDEFPVEVD  
VDNASSEENDSDALDDSDDDLASDDYSDSDSKSGSRKQNKWFKFFGSLDLSIQ  
INEROROMCPACONGGALDWMYHLPLHARTRKARVYKHLRELALEVEKDLQMRG  
ASVTPCEIYVCOMKGLGDEKDYETIWPMPVITIMNTRLDKDNDKMTGLMGQKRS  
DKYALAPRHSYGPQGRGMSVLMFESSATGYLEAERLHELMEGLDRLAMQKRS  
FSGVRLDYGLATKQDLDFNHSQGRTRLEKLSYQEMVKEALQISPDNQLMY  
PKNLSKQNRKARVLESLEIMSEKRLRTLEDNTVQRTMQHSENRDEADRF  
MDSIKQIHREDAENFEMLOQERAKVVGQOQONINPSNDCKRRAREVSSIF  
QEKMEEFVEREEMWIKQEKMEKMKRHIEIFDLEKEFDEALQIMYHGHINED  
D"

1347 CAGATTAATGAACACAGAGGAGTGTCTTCCAGCTTGTCAAGACGACCTGTGTC 1406

Db 743 CAGATAAATGAAACCAAGAGGAGGAGTTCAGCTTGTGCAAGACGACTGTGCTC 802  
QY 1407 ATGATTTGGATATACCTGCACCTCTTACTAGCTCATGCGAGGACAAAGAGGCTAGGCGA 1466  
Db 803 ATGATTTGGATATACCTGCACCTCTTACTAGCTCATGCGAGGACAAAGAGGCTAGGCGA 862  
QY 1467 GTTAACCTCATAGAGAAATTTGGCTGAAGTTTAAAGAAAGATCTAGAGATGAGGCGCA 1526  
Db 863 GTTAACCTCATAGAGAAATTTGGCTGAAGTTTAAAGAAAGATCTAGAGATGAGGCGCA 922  
QY 1527 TCTGTCAATCTTGTGCTGAGATTTATGCGAGCTGAGAGGCTTTGGGTGAGATGAAG 1586  
Db 923 TCTGTCAATCTTGTGCTGAGATTTATGCGAGCTGAGAGGCTTTGGGTGAGATGAAG 982  
QY 1587 GATTAGAAATTTGCTGCGCTCCAAATGTCATCATGTAATCTAGCTGATGAAGGAC 1646  
Db 983 GATTAGAAATTTGCTGCGCTCCAAATGTCATCATGTAATCTAGCTGATGAAGGAC 1042  
QY 1647 GATAACGATTAAG 1658  
Db 1043 GATAACGATTAAG 1054

RESULT 5  
AX078761 1878 bp DNA linear PAT 22-FEB-2001  
LOCUS AX078761  
DEFINITION Sequence 2 from Patent WO0105951.  
ACCESSION AX078761  
VERSION AX078761.1 GI:13158380  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsi.  
1  
Beclin, C., Elmayer, T. and Vaucheret, H.  
Novel sggs plant gene and use thereof  
Patent: WO 0105951-A 2 25-JAN-2001  
AVANTIS CROPS SCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE  
AGRONOMIQUE (FR)

FEATURES  
source  
Location/Qualifiers  
1.1878  
/organism="Arabidopsis thaliana"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:3702"  
1.1878  
/note="unassigned protein product"  
/codon\_start=1  
/protein\_id="CAC32419.1"  
/db\_xref="GI:13158381"  
/translation="MSSRAGPMSKKNVQGYRPEVQHVQGLASODDGESE  
EVSKNKKNKPGNTSGKTWSONSNPPRAGQOQGRGVSGRGNVSGRGNGRG  
IOANISGRGALSRLKTYDNNFVAPVSRPLEGAMWQAGCAHATAVOEPDVEDD  
VONASREENDSALDSDDDLDSDVDSDVSKSHSRKONKFKKPFSLDSTED  
INPQKMHCPACQNPACAIIDRYNHLPLAHARTKARVYKHLRELAIVLENDLQNRG  
ASVTPCGEITYGQKGLGEDEKDYEIWPPEVITIMNTRLDNDKMLGMDNGLSEIF  
DKTEALRARSYVPOGRHGSVLMFESSATGYLEARLHRELAEMGLDIRAMQKXSM  
FSGGVQLVFLATKODLDIENQSOQKRLKELKSYOMVYVEKLESDNQALY  
FKKLSKONHAKVLEESLEINSEKLRTEENRIYVORTKQHEONREEMDAHEFF  
MOSIKOIHERRDAKENFEMLOOQERAKVVGQOQONINPSNDGCRKRAEYVSFTFF  
QEKEMEFVEBERMLIKDEKRMEDMKRHHETFLDEKFEPLALQLMVKHGLHNE  
D"

## ORIGIN

Query Match 29.4%; Score 963; DB 6; Length 1878;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 696 ATAGATTCTAGGCGCTGCTCAATGTTCTAGAGAAAGAAAGCTTACGGGTGTTATAGGCTT 755  
1 ATAGATTCTAGGCGCTGCTCAATGTTCTAGAGAAAGAAAGCTTACGGGTGTTATAGGCTT 60

QY 756 GAGTTGAACAGTTGGTTCAAGTTTGGAGGAGACGAGACTGGCTTCTTCAAGATGAT 815  
Db 61 GAGTTGAACAGTTGGTTCAAGTTTGGAGGAGACGAGACTGGCTTCTTCAAGATGAT 120  
QY 816 GGAGAGAGTGGAGAGTCAATTTCCAAAGAAACAAGAACAAACAGAGAAACCTTTGGA 875  
Db 121 GGAGAGAGTGGAGAGTCAATTTCCAAAGAAACAAGAACAAACAGAGAAACCTTTGGA 180  
QY 876 AAACTTGGGTTTCTCAAGATTTGAAATCTCTTGAAGCTTTGGGTGCTGAGAGAGG 935  
Db 181 AAACTTGGGTTTCTCAAGATTTGAAATCTCTTGAAGCTTTGGGTGCTGAGAGAGG 240  
QY 936 AGAGTAGACAGTATCTGGAGAGAGAAACAATGTATCCGAGAGAGTAAACGCAATGCT 995  
Db 241 AGAGTAGACAGTATCTGGAGAGAGAAACAATGTATCCGAGAGAGTAAACGCAATGCT 300  
QY 996 CGGGGCAATCAAGCTAACATATCTGCTCGGGAGAGAGCTTGAAGAGAAATATGATAAC 1055  
Db 301 CGGGGCAATCAAGCTAACATATCTGCTCGGGAGAGAGCTTGAAGAGAAATATGATAAC 360  
QY 1056 AACTTTGTGCAACCCCACTGTATCTCGCTCTCTTGAAGAGAGATGAAATTTGGCAG 1115  
Db 361 AACTTTGTGCAACCCCACTGTATCTCGCTCTCTTGAAGAGAGATGAAATTTGGCAG 420  
QY 1116 GCAAGAGAGGTTTCTGCTGAGACACAGCTGTGAGAGAGTTTCTGAGCTGAGAGATGAT 1175  
Db 421 GCAAGAGAGGTTTCTGCTGAGACACAGCTGTGAGAGAGTTTCTGAGCTGAGAGATGAT 480  
QY 1176 GTGATATATGCTTCTGAGAGAGAGATGATTCGATGCTTTGGATGATTTCTGATGACAG 1235  
Db 481 GTGATATATGCTTCTGAGAGAGAGATGATTCGATGCTTTGGATGATTTCTGATGACAG 540  
QY 1236 CTTCGAGAGATGATTTATGATCTGATGATGATGATGATGATGATGATGATGATGATG 1295  
Db 541 CTTCGAGAGATGATTTATGATCTGATGATGATGATGATGATGATGATGATGATGATG 600  
QY 1296 AATTAAGTGTGTTCAAAAAGTTCTTTGGCAGCTTGATGATGATGATGATGATGATGAT 1355  
Db 601 AATTAAGTGTGTTCAAAAAGTTCTTTGGCAGCTTGATGATGATGATGATGATGATGAT 660  
QY 1356 GAACCAAGAGAGAGAGTGTGCTGATGCTTGCAGAACGAGCTGTGCTCATGATGATG 1415  
Db 661 GAACCAAGAGAGAGAGTGTGCTGATGCTTGCAGAACGAGCTGTGCTCATGATGATG 720  
QY 1416 TATAACCTGACCTCTTACTAGCTCATGCGAGAGACAAAGAGACTRAGGAGTTAAGCTC 1475  
Db 721 TATAACCTGACCTCTTACTAGCTCATGCGAGAGACAAAGAGACTRAGGAGTTAAGCTC 780  
QY 1476 CATAGAGATTTGGCTGAATTTTAAAGAAAGATCTACAGATGAGAGAGGCTATCTGAT 1535  
Db 781 CATAGAGATTTGGCTGAATTTTAAAGAAAGATCTACAGATGAGAGAGGCTATCTGAT 840  
QY 1536 CTTTGTGTGAGATTTATGAGAGAGTGGTGGTGGTGAAGATGAAGATTAATGA 1595  
Db 841 CTTTGTGTGAGATTTATGAGAGAGTGGTGGTGGTGAAGATGAAGATTAATGA 900  
QY 1596 ATTGTGTGCTGCTCAATGCTCATCATGAAATCTAGACTGATTAAGACGATTAACGAT 1655  
Db 901 ATTGTGTGCTGCTCAATGCTCATCATGAAATCTAGACTGATTAAGACGATTAACGAT 960  
QY 1656 AAG 1658  
Db 961 AAG 963

RESULT 6  
BT004380  
LOCUS BT004380 1909 bp mRNA linear PLN 14-FEB-2003  
DEFINITION Arabidopsis thaliana clone U20243 unknown protein (At5g23570) mRNA,  
complete cds.  
ACCESSION BT004380  
VERSION BT004380.1 GI:28393932

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
COMMENT  
JOURNAL  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
COMMENT

FLI CDNA.  
Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 1909)  
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,  
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,  
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,  
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,  
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,  
Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,  
Ecker, J.R. and Theologis, A.  
Arabidopsis Open Reading Frame (ORF) Clones  
Unpublished  
2 (bases 1 to 1909)  
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,  
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,  
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,  
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,  
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,  
Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,  
Ecker, J.R. and Theologis, A.  
Direct Submission  
Submitted (14-FEB-2003) Plant Gene Expression Center, 800 Buchanan  
Street, Albany, CA 94710, USA  
The RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAPL cDNAs (RAPL CDNA: 'RIKEN  
Arabidopsis Full-length cDNA': Seki, M., Narusaka, M., Ishida, J.,  
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y. and Theologis, A.  
The Salk, Stanford, PGEC (SSP) Consortium members constructed and  
sequenced the PUNI (ORF) clones using the RAPL cDNAs: Yamada, K.,  
Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,  
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,  
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,  
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,  
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,  
Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,  
Ecker, J.R. and Theologis, A.  
Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally  
to this work. Shinzaki, K. (RIKEN GSC) and Theologis, A. (SSP  
/PGEC) contributed equally to this work as PIs.  
Annotation based on July 2002 version of the Arabidopsis genome  
submitted to Genbank.  
Location/Qualifiers  
1..1909  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="U20243"  
/ecotype="Columbia"  
/note="This clone is in PUNI 51."  
1..1909  
/gene="At5g23570"  
1..1878  
/gene="At5g23570"  
/codon\_start=1  
/evidence="experimental"  
/product="unknown protein"  
/protein\_id="AA042374.1"  
/db\_xref="GI:28393933"  
/translation="MSSRAGPMSKKNVGGYRPEVQIYVGLAGTRLLASDODGEM  
EYISKKNKPKGNTSGKTVVSONSPRAAGGQOQSGNSVGRNNVSGGNGRG  
IVANISGRGALSRKYDNNVAPPPVSRPPLGGKMMWQARGSAAHTVQEPVEDD  
VDNAISENDSDALDDSDDLASDDYDSDVSOQSGSRKQNKWFKFPGSDLSIED  
INPBRQHPACONGPAIDWYNLHPLAHARTGARVRLKWFLEKDLQMRG  
ASTYPCGEIYQWKGIGDEKQDEIVPMPVIMNTRLDKNDNKLAEGLDNGDELIEYF  
DKTEALRAHRSYSPQGRGMSVILMFESSATYLAERLRELAEGMDLQIANGQKRM  
PSSGVROLYGFATKODLDIFNDSOGKTRKLEKSYQEMVYVKELRQISENQQLNY

ORIGIN  
3' UTR  
Query Match 29.4%; Score 963; DB 8; Length 1909;  
Best Local Similarity 100.0%; Pred.No. 0; Mismatches 0; Gaps 0;  
Matches 963; Conservative 0; Indels 0; Gaps 0;  
D  
/gene="At5g23570"  
1879..1909  
PNNKSKONKAKVLEESLEIWESEKTRTAEDNRIVRQRTKQHEQNEENDADHRRF  
MSIQIHERPRDAXENEMILQOGERARVINGOONINPSSNDCKRAEVSFIEF  
QEKEMEEFVEEREMILIKDEKEMDMKKRHHEIFDLKEPDEALEQIMYKHGLNED

696 ATGAGTTCTAGGCGTGGTCCAAATGTCTAAGAAAAGAACCTTCAGGCTGTTATAGGCT 755  
1 ATGAGTTCTAGGCGTGGTCCAAATGTCTAAGAAAAGAACCTTCAGGCTGTTATAGGCT 60  
756 GAGGTTGAACAGTTGGTTCAAGTTTGGCGAGGACGAGCTGGCTCTTCACAAATGAT 815  
61 GAGGTTGAACAGTTGGTTCAAGTTTGGCGAGGACGAGCTGGCTCTTCACAAATGAT 120  
816 GGAGGAGAGTGGAGGAGTCAATTTCCAGAGAAACAGAACCAAGCAAACTTCTGGA 875  
121 GGAGGAGAGTGGAGGAGTCAATTTCCAGAGAAACAGAACCAAGCAAACTTCTGGA 180  
876 AAAAATTGGGTTTCTCAGAATTCGATCTCTAGAGCTTGGGCTGTCAGCAAGAGG 935  
181 AAAAATTGGGTTTCTCAGAATTCGATCTCTAGAGCTTGGGCTGTCAGCAAGAGG 240  
936 AGAGTTAGCAAGTATCTGGGAGAGAAACATATATCCGGAGAGGTTAACCGCAATGT 995  
241 AGAGTTAGCAAGTATCTGGGAGAGAAACATATATCCGGAGAGGTTAACCGCAATGT 300  
996 CGGGGCAATTCAGCTAACATATCTGGGAGAGAGAGGAGGAGGAGGAGGAGGAGG 1055  
301 CGGGGCAATTCAGCTAACATATCTGGGAGAGAGAGGAGGAGGAGGAGGAGGAGG 360  
1056 AACTTTTGGGAGACCCCACTGATCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1115  
361 AACTTTTGGGAGACCCCACTGATCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420  
1116 GGAAGAGAGGTTTGGCTCAGACACACAGCTGACAGAGTTTCTGACGTGGAGATGAT 1175  
421 GGAAGAGAGGTTTGGCTCAGACACACAGCTGACAGAGTTTCTGACGTGGAGATGAT 480  
1176 GTGGATATCTTCTGAGAGAGAGATGATTCGATGATGATGATGATGATGATGATGAT 1235  
481 GTGGATATCTTCTGAGAGAGAGATGATTCGATGATGATGATGATGATGATGATGAT 540  
1236 CTGGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1295  
541 CTGGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
1296 AATAAGTGTTCAAAAAGTCTTGGGAGGCTGATGATGATGATGATGATGATGATGAT 1355  
601 AATAAGTGTTCAAAAAGTCTTGGGAGGCTGATGATGATGATGATGATGATGATGATGAT 660  
1356 GAACCAAGAGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1415  
661 GAACCAAGAGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
1416 TATTAACCTGACCTCTACTAGCTCATGCGAGACAAAAGAGCTTAGGCTTAAGCTC 1475  
721 TATTAACCTGACCTCTACTAGCTCATGCGAGACAAAAGAGCTTAGGCTTAAGCTC 780  
1476 CATAGAGATTTGGCTGAAGTTTGAAGAGAGATCTACAGATGAGAGGCGCATCTGCAAT 1535  
781 CATAGAGATTTGGCTGAAGTTTGAAGAGAGATCTACAGATGAGAGGCGCATCTGCAAT 840  
1536 CCTGTGTGATGATTTATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1595  
841 CCTGTGTGATGATTTATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900  
1596 ATTGTGGGCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1655

|||||  
Db 901 ATTGTCGCGCTCCATCATGATCATCATGAATAGACTGATGAGACGATAACGAT 960  
QY 1656 AAG 1658  
|||  
Db 961 AAG 963  
|||  
RESULT 7  
ATH528171  
LOCUS 650 bp DNA linear PLN 29-MAR-2003  
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone  
157C08.  
ACCESSION AJ528171 GI:26796431  
VERSION 1  
KEYWORDS left border; T-DNA flanking sequence;  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
REFERENCE 1  
AUTHORS Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F.,  
Chauvin, S., Bechold, N., Crenaud, C., Derose, R., Pelletier, G.,  
Lepiniec, L., Caboche, M. and Lecharny, A.  
TITLE T-DNA integration into the Arabidopsis genome depends on sequences  
of pre-insertion sites  
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)  
MEDLINE 22363535  
PUBMED 12446565  
REFERENCE 2 (bases 1 to 650)  
AUTHORS Balzerque, S.  
TITLE Direct Submission  
JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue  
Gaston Cremieux, 91057 Evry cedex, FRANCE  
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana  
plants from INRA (Versailles). The DNA fragment(s) resulting from  
the PCR were directly sequenced from the left or the right border  
to determine the genomic sequence flanking the insertion. T-DNA  
derived sequences were removed. Information to order the  
corresponding mutant line and a link to a database providing a  
graphical display of the insertion site are available at  
http://dbsgap.versailles.inra.fr/publiclines/. This sequence has  
been generated in the framework of the French plant genomics  
program 'Genoplante' (http://www.genoplante.com and  
http://genoplante-info.infobiogen.fr).  
FEATURES  
source  
1. 650  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/cultivar="Wassaillewjka"  
/db\_xref="taxon:3702"  
/clone="157C08"  
misc\_feature  
1. 650  
/note="T-DNA flanking sequence  
left border"  
ORIGIN  
Query Match 16.5%; Score 541; DB 8; Length 650;  
Best Local Similarity 99.8%; Pred. No. 1.9e-296;  
Matches 591; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1428 CCTCTACTGCTCATCGAGACAAAAGAGCTAGCGGAGTTAGCTCCATAGAGATTG 1487  
Db 1 CCTCTACTGCTCATCGAGACAAAAGAGCTAGCGGAGTTAGCTCCATAGAGATTG 60  
QY 1488 GCTGAAGTTTGAAGAAGATCTACAGATGAGAGGGGAGTCTCTCTCTGTTGAG 1547  
Db 61 GCTGAAGTTTGAAGAAGATCTACAGATGAGAGGGGAGTCTCTCTCTGTTGAG 120  
QY 1548 ATTATGCGCAGTGAAGGGTTGGTGAAGATGAAGAAGATTATGCTCTGCGCT 1607  
|||||

|||||  
Db 121 ATTATGCGCAGTGAAGGGTTGGTGAAGATGAAGAAGATTATGCTCTGCGCT 180  
QY 1608 CCAATGTCATCATGATGAATCTAGACTGATTAAGACGATTAAGATGAGATTC 1667  
|||  
Db 181 CCAATGTCATCATGATGAATCTAGACTGATTAAGACGATTAAGATGAGATTC 240  
QY 1668 TTCTGCTTTTAACTCTTTAAATTTTCTTTCATCTTCTGATCTTGAAGTTACAT 1727  
Db 241 TTCTGCTTTTAACTCTTTAAATTTTCTTTCATCTTCTGATCTTGAAGTTACAT 300  
QY 1728 GTATGCGCTCGGACATGCGCAACCAAGAGCTGCTGGAATCTTCCACAGTATGAGGCTCT 1787  
Db 301 GTATGCGCTCGGACATGCGCAACCAAGAGCTGCTGGAATCTTCCACAGTATGAGGCTCT 360  
QY 1788 TAGAGAGCGCATTCCTTATGCTCAGAGGCGCATGCTGGAGATGATGTTCTGATGTTGA 1847  
Db 361 TAGAGAGCGCATTCCTTATGCTCAGAGGCGCATGCTGGAGATGATGTTCTGATGTTGA 420  
QY 1848 GAGCAGTGCCTGCTATTTTGAAGCGCAACGCTCCACCGGAGTTAGCTGAGATGGG 1907  
Db 421 GAGCAGTGCCTGCTATTTTGAAGCGCAACGCTCCACCGGAGTTAGCTGAGATGGG 480  
QY 1908 GTTAAATGAAATTCCTGCGGCTCAGAGCGCATGATGTTTCTGAGAGTTCGCCACT 1967  
Db 481 GTTAAATGAAATTCCTGCGGCTCAGAGCGCATGATGTTTCTGAGAGTTCGCCACT 540  
QY 1968 GTATGCGCTCTTGCAGCAAGATCTGAGCATTTCAATCAACTCT 2019  
Db 541 GTATGCGCTCTTGCAGCAAGATCTGAGCATTTCAATCAACTCT 592  
|||  
RESULT 8  
ATH527741  
LOCUS 45 bp DNA linear PLN 29-MAR-2003  
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone  
146F03.  
ACCESSION AJ527741 GI:26796001  
VERSION 1  
KEYWORDS left border; T-DNA flanking sequence.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
REFERENCE 1  
AUTHORS Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F.,  
Chauvin, S., Bechold, N., Crenaud, C., Derose, R., Pelletier, G.,  
Lepiniec, L., Caboche, M. and Lecharny, A.  
TITLE T-DNA integration into the Arabidopsis genome depends on sequences  
of pre-insertion sites  
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)  
MEDLINE 22363535  
PUBMED 12446565  
REFERENCE 2 (bases 1 to 45)  
AUTHORS Balzerque, S.  
TITLE Direct Submission  
JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue  
Gaston Cremieux, 91057 Evry cedex, FRANCE  
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana  
plants from INRA (Versailles). The DNA fragment(s) resulting from  
the PCR were directly sequenced from the left or the right border  
to determine the genomic sequence flanking the insertion. T-DNA  
derived sequences were removed. Information to order the  
corresponding mutant line and a link to a database providing a  
graphical display of the insertion site are available at  
http://dbsgap.versailles.inra.fr/publiclines/. This sequence has  
been generated in the framework of the French plant genomics  
program 'Genoplante' (http://www.genoplante.com and  
http://genoplante-info.infobiogen.fr).  
FEATURES  
source  
1. 45  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"



Query Match 0.7%; Score 24; DB 2; Length 170765;  
 Best Local Similarity 100.0%; Pred. No. 0.51;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ACAAACAAACAAATTAAGCAAG 25  
 Db 65301 ACAAACAAACAAATTAAGCAAG 65278

RESULT 10  
 AF057056/c  
 LOCUS  
 DEFINITION Uroleucon erigeronensis NADH dehydrogenase subunit 1-like protein (ND1) gene, partial sequence; mitochondrial gene for mitochondrial product.  
 ACCESSION AF057056  
 VERSION AF057056.1 GI:8745205  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macrosiphini; Uroleucon.  
 REFERENCE  
 AUTHORS Moran,N.A., Kaplan,M.E., Gelsey,M.J., Murphy,T.G. and Scholes,E.A.  
 TITLE Phylogenetics and evolution of the aphid genus Uroleucon based on mitochondrial and nuclear DNA sequences  
 JOURNAL Syst. Entomol. 24 (1), 85-93 (1999)  
 REFERENCE 2 (bases 1 to 554)  
 AUTHORS Moran,N.A., Kaplan,M.E., Gelsey,M.J., Murphy,T.G. and Scholes,E.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-APR-1998) Ecology and Evolutionary Biology, University of Arizona, Biological Sciences West, Tucson, Arizona 85721, USA

FEATURES  
 source 1..554  
 /organism="Uroleucon erigeronensis"  
 /organelle="mitochondrion"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:87314"  
 <1..>554  
 /gene="ND1"  
 1..554  
 /gene="ND1"  
 /note="similar to NADH dehydrogenase subunit 1"

ORIGIN  
 Query Match 0.7%; Score 23; DB 3; Length 554;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3151 GTTTTGTGTAATTCAT 3173  
 Db 468 GTTTTGTGTAATTCAT 446

RESULT 11  
 BX004860  
 LOCUS  
 DEFINITION Dario rerio clone BUSM1-144B24, WORKING DRAFT SEQUENCE, 4 unordered pieces.  
 ACCESSION BX004860  
 VERSION BX004860.1 GI:26985650  
 KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.  
 SOURCE Dario rerio (zebrafish)  
 ORGANISM Dario rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
 REFERENCE 1 (bases 1 to 101080)  
 AUTHORS Burdon,J.  
 TITLE Direct Submission

JOURNAL  
 Submitted (10-DEC-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk  
 Clone requests: clonerequest@sanger.ac.uk  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: zface@sanger.ac.uk  
 ----- Project Information  
 Center project name: dz144B24  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: Dye-terminator; 100% of reads  
 Consensus quality: 99769 bases at least Q40  
 Consensus quality: 100257 bases at least Q30  
 Consensus quality: 100510 bases at least Q20  
 Insert size: 100780; sum-of-contigs  
 Insert size: 107197; 9.7% error; agarose-fp  
 Quality coverage: 5.75x in Q20 bases; sum-of-contigs Quality coverage: 5.56x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. \* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
 \* 1 34070: contig of 34070 bp in length  
 \* 34071 34170: gap of 100 bp  
 \* 34171 67630: contig of 33460 bp in length  
 \* 67631 67730: gap of 100 bp  
 \* 67731 81714: contig of 13984 bp in length  
 \* 81715 81814: gap of 100 bp  
 \* 81815 101080: contig of 19266 bp in length.

FEATURES  
 source 1..101080  
 /organism="Dario rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="BUSM1-144B24"  
 /clone\_1b="BUSM1"  
 1..34070  
 /note="assembly\_fragment:00417  
 fragment\_chain:1  
 clone\_end:77  
 vector\_side:left"  
 34171..67630  
 /note="assembly\_fragment:00598  
 fragment\_chain:1"  
 67731..81714  
 /note="assembly\_fragment:00262  
 fragment\_chain:1"  
 81815..101080  
 /note="assembly\_fragment:00119  
 fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:right"

ORIGIN  
 Query Match 0.7%; Score 23; DB 2; Length 101080;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 96 AATGAAAAAATCAACTGA 118  
 Db 39791 AATGAAAAAATCAACTGA 39769

RESULT 12  
 AL845510/c  
 LOCUS  
 AL845510 101241 bp DNA linear VRT 30-JAN-2003



DEFINITION Zebrafish DNA sequence from clone BUSM1-132M23 in linkage group 7, complete sequence.

ACCESSION AL845510

VERSION AL845510.8 GI:28172206

KEYWORDS HTG.

SOURCE Dario rerio (zebrafish)

ORGANISM Dario rerio

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Dario.

AUTHORS Bates, K.

TITLE Direct Submission

JOURNAL Submitted (29-JAN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk

COMMENT On Jan 30, 2003 this sequence version replaced gi:27801659.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfish-help@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: BM, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)

Repeats identified by the Repeat discovery system (Zhong Bao and Sean Eddy, submitted), and those beginning 'dr' (were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see [http://www/Projects/D\\_rerio/fishmark.shtml](http://www/Projects/D_rerio/fishmark.shtml)

BUSM1-132M23 is from a Zebrafish PAC library

VECTOR: pCYPAC-6.

FEATURES

source

1..101241

/organism="Dario rerio"

/mol\_type="genomic DNA"

/db\_xref="taxon:7955"

/clone="BUSM1-132M23"

/clone\_1lb="BUSM1"

ORIGIN

Query Match 0.7%; Score 23; DB 5; Length 101241;

Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 AATGAAAACAAATCAACTGA 118

DB 39751 AATGAAAACAAATCAACTGA 39729

RESULT 13

AC107028/c 126716 bp DNA linear PRI 29-MAR-2003

LOCUS AC107028 Homo sapiens 3 BAC RP11-547K2 (Roswell Park Cancer Institute Human BAC library) complete sequence.

DEFINITION BAC RP11-547K2 (Roswell Park Cancer Institute Human BAC library) complete sequence.

ACCESSION AC107028

VERSION AC107028.5 GI:28006935

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banks, T., Barbieri, J., Benton, J., Bimble, K., Blankensburg, K., Bonin, D., Bouch, J., Bowie, S., Brice, M., Brown, E., Brown, M., Bryant, N.P., Buay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homel, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, J.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, M., Louissaged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, J., Mitchell, T., Monabhat, K., Moore, S., Morgan, M., Morris, T., Morris, S., Moser, M., Neal, D., Nicolson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nicholson, E., Norken, S., Ogburn, M., Okunoda, G., Oregano, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Qulies, M., Ren, Y., Rivers, M., Rojals, A., Rojudo, K., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitani, N., Stinson, I., Sodergren, S., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taylor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Woodson, S., Worley, K., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstein, G., and Gibbs, R.

REFERENCE Direct Submission

JOURNAL Unpublished

TITLE 2 (bases 1 to 126716)

REFERENCE Worley, K.C.

AUTHORS Direct Submission

TITLE Submitted (14-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 126716)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (29-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 126716)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (30-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 5 (bases 1 to 126716)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (02-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA



REFERENCE 6 (bases 1 to 126716)  
 AUTHORS Worley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 7 (bases 1 to 126716)  
 AUTHORS Worley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-JAN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 8 (bases 1 to 126716)  
 AUTHORS Worley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 9 (bases 1 to 126716)  
 AUTHORS Worley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jan 29, 2003 this sequence version replaced gi:20340431.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:  
 STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.  
 Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) For Human and Mouse sequences.  
 Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/gcnbank.annotation.html>.

FEATURES  
 source Location/Qualifiers  
 1..126716  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /clone="RP11-547K2"  
 complement(1..3011)  
 /note="overlaps bases 1..3011 of clone ACO26305"  
 /function="clone overlap"  
 complement(28..209)  
 repeat\_region  
 misc\_feature  
 repeat\_region

repeat\_region /rpt\_family="MER34"  
 219..973  
 /rpt\_family="L1M4"  
 repeat\_region 1409..1455  
 /rpt\_family="AluY"  
 repeat\_region 1456..1754  
 /rpt\_family="AluY"  
 repeat\_region 1762..1998  
 /rpt\_family="AluSc"  
 repeat\_region 2145..2195  
 /rpt\_family="AT\_rich"  
 repeat\_region 2612..2921  
 /rpt\_family="AluSx"  
 repeat\_region 3493..3520  
 /rpt\_family="AT\_rich"  
 repeat\_region 3573..3598  
 /rpt\_family="AT\_rich"  
 repeat\_region 5930..5955  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(6168..6476)  
 /rpt\_family="AluSp"  
 repeat\_region 6945..6984  
 /rpt\_family="(TTTA)n"  
 repeat\_region complement(6986..7276)  
 /rpt\_family="AluSx"  
 repeat\_region 11796..11975  
 /rpt\_family="MER5A"  
 repeat\_region complement(13903..13967)  
 /rpt\_family="L2"  
 repeat\_region 14192..14235  
 /rpt\_family="MIR"  
 repeat\_region 16052..16078  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(17082..17461)  
 /rpt\_family="WSTD"  
 repeat\_region complement(17583..17934)  
 /rpt\_family="MER21B"  
 repeat\_region 17944..17997  
 /rpt\_family="(TTATA)n"  
 repeat\_region complement(18003..18257)  
 /rpt\_family="AluDb"  
 repeat\_region complement(18261..18557)  
 /rpt\_family="MER21B"  
 repeat\_region 19148..19215  
 /rpt\_family="AT\_rich"  
 repeat\_region 19234..19342  
 /rpt\_family="FLAM\_C"  
 repeat\_region 19377..19428  
 /rpt\_family="A-rich"

Query Match 0.7%; Score 23; DB 9; Length 126716;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 23; Conservative 0; Mismatches 0;  
 Oy 1673 TCTTTACTCTTTATTTTCCT 1695  
 Db 47492 TCTTTACTCTTTATTTTCCT 47470

RESULT 14  
 LOCUS BX005128/c  
 DEFINITION zebrafish DNA sequence from clone CH211-129H4 in linkage group 17,  
 complete sequence.  
 ACCESSION BX005128  
 VERSION BX005128.4 GI:37518236  
 KEYWORDS HTG.  
 SOURCE Dario rerio (zebrafish)  
 ORGANISM Dario rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Dario.  
 REFERENCE 1 (bases 1 to 152345)

AUTHORS  
TITLE  
JOURNAL  
COMMENT

## COMMENT

Sehra, H.  
Direct Submission  
Submitted (03-Oct-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Oct 5 2003 this sequence version replaced g1:3014105.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30), an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
EM, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information  
on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Repeat names  
beginning 'Dr' were identified by the Recon repeat discovery system  
(Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dir'  
were identified by Rick Waterman (Stephen Johnson lab, WashU). For  
further information see  
[http://www.sanger.ac.uk/Projects/D\\_reio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_reio/fishmask.shtml)  
CH211-129H14 is from a CHORI-211 BAC library

VECTOR: pTRABAC2.1  
Cloned derived zebrafish pUC subclones occasionally display  
inconsistency over the length of mononucleotide A/T runs and  
conserved TA repeats. Where this is found the longest good quality  
representation will be submitted.  
Location/Qualifiers

1. 152345  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="CH211-129H14"  
/clone\_1ib="CHORI-211"

## ORIGIN

Query Match 0.7%; Score 23; DB 5; Length 152345;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

3141 TTACTATTTTGTCTTTTGTCT 3163  
2054 TTAGTATTTTGTCTTTTGTCT 2032

RESULT 15  
AC048381 180925 bp DNA linear HTG 24-ANG-2002  
LOCUS Homo sapiens chromosome 3 clone RP11-547K2 map 3, WORKING DRAFT  
DEFINITION  
SEQUENCE, 37 unordered pieces.  
AC048381  
AC048381.3 GI:8084282  
VERSION  
KEYWORDS HTGS PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
1 (bases 1 to 180925)

AUTHORS  
TITLE  
JOURNAL  
REFERENCE

Bliren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens chromosome 3, clone RP11-547K2  
Unpublished  
2 (bases 1 to 180925)  
Bliren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Baettlen, V., Bede, F.,  
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,  
Campomiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collins, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Domingo, M., Doyle, M., Ferrer, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Lacombe, K., Lamazeres, R., Landers, T., Lenockky, J.,  
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McKernan, K., McPherson, R.,  
Meldrum, J., Menus, L., Minova, T., Miranda, C., Miñana, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Olliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Submitted (14-Apr-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 180925)

REFERENCE  
AUTHORSTITLE  
JOURNAL  
COMMENT

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced g1:770638.  
All repeats were identified using RepeatMasker:  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: W1BR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center Project name: 547\_K\_2  
Center clone name: 547\_K\_2  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 158029 bases at least Q40  
Consensus quality: 168695 bases at least Q30  
Consensus quality: 173285 bases at least Q20  
Insert size: 183000; agarose-1p

Insert size: 177325; sum-of-contrigs  
Quality coverage: 3.4 in Q20 bases; agarose-fp  
Quality coverage: 3.6 in Q20 bases; sum-of-contrigs

NOTE: This is a 'working draft' sequence. It currently consists of 37 contrigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contrigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1010: contrig of 1010 bp in length
1011 1110: gap of 100 bp
1111 2487: contrig of 1377 bp in length
2488 2587: gap of 100 bp
2588 3720: contrig of 1133 bp in length
3721 3820: gap of 100 bp
3821 4892: contrig of 1072 bp in length
4893 4992: gap of 100 bp
4993 5489: contrig of 497 bp in length
5490 5589: gap of 100 bp
5590 7151: contrig of 1562 bp in length
7152 7251: gap of 100 bp
7252 8655: contrig of 1404 bp in length
8656 8755: gap of 100 bp
8756 10134: contrig of 1379 bp in length
10135 10234: gap of 100 bp
10235 13107: contrig of 2873 bp in length
13108 13207: gap of 100 bp
13208 15265: contrig of 2058 bp in length
15266 15365: gap of 100 bp
15366 17768: contrig of 2403 bp in length
17769 17868: gap of 100 bp
17869 20233: contrig of 2365 bp in length
20234 20333: gap of 100 bp
20334 22968: contrig of 2635 bp in length
22969 23068: gap of 100 bp
23069 25333: contrig of 2265 bp in length
25334 25433: gap of 100 bp
25434 28978: contrig of 3545 bp in length
28979 29078: gap of 100 bp
29079 31604: contrig of 2526 bp in length
31605 31704: gap of 100 bp
31705 34400: contrig of 2696 bp in length
34401 34500: gap of 100 bp
34501 37255: contrig of 2755 bp in length
37256 37355: gap of 100 bp
37356 40057: contrig of 2702 bp in length
40058 40157: gap of 100 bp
40158 44190: contrig of 4033 bp in length
44191 44290: gap of 100 bp
44291 46785: contrig of 2495 bp in length
46786 46885: gap of 100 bp
46886 52597: contrig of 5712 bp in length
52598 52697: gap of 100 bp
52698 59026: contrig of 6329 bp in length
59027 59126: gap of 100 bp
59127 64431: contrig of 5305 bp in length
64432 64531: gap of 100 bp
64532 69115: contrig of 4584 bp in length
69116 69215: gap of 100 bp
69216 73866: contrig of 4651 bp in length
73867 73966: gap of 100 bp
73967 79516: contrig of 5550 bp in length
79517 79616: gap of 100 bp
79617 85688: contrig of 6072 bp in length
85689 85788: gap of 100 bp
85789 92058: contrig of 6270 bp in length
92059 92158: gap of 100 bp
92159 99039: contrig of 6881 bp in length
99040 99139: gap of 100 bp
99140 108885: contrig of 9746 bp in length

```

## FEATURES

## source

```

* 10886 108985: gap of 100 bp.
* 108986 116971: contrig of 7986 bp in length
* 116972 117071: gap of 100 bp
* 117072 126718: contrig of 9647 bp in length
* 126719 126818: gap of 100 bp
* 126819 137354: contrig of 10536 bp in length
* 137355 137454: gap of 100 bp
* 137455 146943: contrig of 9489 bp in length
* 146944 147043: gap of 100 bp
* 147044 157785: contrig of 10742 bp in length
* 157786 157885: gap of 100 bp
* 157886 180925: contrig of 23040 bp in length.

```

## Location/Qualifiers

```

1..180925
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-547K2"
/clone_lib="RPC1-11 Human Male BAC"
1..1010
/note="assembly_fragment"
111..2487
/note="assembly_fragment"
2588..3720
/note="assembly_fragment"
3821..4892
/note="assembly_fragment"
4993..5489
/note="assembly_fragment"
clone_end:T7
vector_side:right"
5590..7151
/note="assembly_fragment"
7252..8655
/note="assembly_fragment"
8756..10134
/note="assembly_fragment"
10235..13107
/note="assembly_fragment"
13208..15265
/note="assembly_fragment"

```

## Query Match

Best Local Similarity 100.0%; Score 23; DB 2; Length 180925;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1673 TCTTTACTTCTTAATTTTCT 1695
Db 16692 TCTTTACTTCTTAATTTTCT 16714

```

Search completed: April 9, 2004, 05:51:24  
Job time : 8467.3 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2004, 16:25:39 ; Search time 837.022 Seconds  
(without alignments)  
16621.841 Million cell updates/sec

Title: US-10-030-829-1  
Perfect score: 3275  
Sequence: 1 gcaacaacaacaataa.....caaatataggggtacaaat 3275

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3275	100.0	3275	4	AAFP25373
2	963	29.4	1878	4	AAFP25374
3	27	0.8	27	4	AAFP25372
4	23	0.7	23	4	AAFP25371
5	22	0.7	6136	6	ABK40041
6	22	0.7	6301	6	ABN80100
7	22	0.7	9888	6	ABL33241
8	21	0.6	6048	6	ABL32509
9	21	0.6	6072	6	ABL32031
10	21	0.6	8245	4	AA546448
11	21	0.6	13453	6	ABN80176
12	21	0.6	52302	8	ADA02738
13	21	0.6	52302	8	ADA02746
14	21	0.6	52302	9	ADC85218
15	20	0.6	420	3	AA52596
16	20	0.6	421	3	AA52596
17	20	0.6	476	6	ABZ78073
18	20	0.6	505	3	AA560012
19	20	0.6	536	6	ABO34916
20	20	0.6	536	6	ABQ34917
21	20	0.6	596	3	AAFI2773
22	20	0.6	627	6	ABO52242
23	20	0.6	627	6	ABO52243

## ALIGNMENTS

C	24	20	0.6	654	2	AA27594	AA27594 Nucleotid
C	25	20	0.6	700	4	AA193017	AA193017 Human inf
C	26	20	0.6	764	6	ABO30109	ABO30109 Oligonuc1
C	27	20	0.6	764	6	ABO30108	ABO30108 Oligonuc1
C	28	20	0.6	843	6	ABK48300	ABK48300 Human inh
C	29	20	0.6	928	9	ADD47853	ADD47853 Rat gene
C	30	20	0.6	928	9	ADD47852	ADD47852 Rat gene
C	31	20	0.6	928	9	ADD47852	ADD47852 Rat gene
C	32	20	0.6	1023	6	ABT07654	ABT07654 Human bre
C	33	20	0.6	1106	4	AA125105	AA125105 Human bre
C	34	20	0.6	1461	7	ACA23079	ACA23079 Prokaryot
C	35	20	0.6	1876	5	ABA20651	ABA20651 Human ner
C	36	20	0.6	2501	9	ADB54297	ADB54297 Pretreat
C	37	20	0.6	2501	9	ADB54169	ADB54169 Pretreat
C	38	20	0.6	3077	7	ABZ10070	ABZ10070 Haematopo
C	39	20	0.6	3077	7	ABZ10070	ABZ10070 Haematopo
C	40	20	0.6	5278	4	AA546376	AA546376 Tumour su
C	41	20	0.6	5278	6	ABL32823	ABL32823 Human imm
C	42	20	0.6	5376	6	ABK34150	ABK34150 Human imm
C	43	20	0.6	5453	6	ABK31481	ABK31481 Signal tr
C	44	20	0.6	5453	6	ABL70450	ABL70450 Chemical1
C	45	20	0.6	5453	6	AA561413	AA561413 Human gen

RESULT 1	AAFP25373	AAFP25373 standard; DNA; 3275 BP.
ID	AAFP25373	
AC	AAFP25373	
XX	15-MAY-2001	(first entry)
DT		
XX		
DE		Genomic sequence of the Arabidopsis SGS3 gene.
XX		
KM		SGS3 gene; post-transcriptional inactivation; RNA degradation;
KW		viral resistance; resistance; fatty acid content; protein content; ss.
XX		
OS		Arabidopsis thaliana.
XX		
FH	Key	Location/Qualifiers
FT	exon	696..1658
FT		/*tag= a
FT	intron	1659..1731
FT		/*tag= b
FT	exon	1732..2023
FT		/*tag= c
FT	intron	2024..2134
FT		/*tag= d
FT	exon	2135..2379
FT		/*tag= e
FT	intron	2380..2481
FT		/*tag= f
FT	exon	2482..2648
FT		/*tag= g
FT	intron	2649..2738
FT		/*tag= h
FT	exon	2739..2949
FT		/*tag= i
XX		
PN	MO200105951-A2.	
XX		
PD	25-JAN-2001.	
XX		
PF	13-JUL-2000; 2000WO-FR002052.	
XX		
PR	16-JUL-1999; 99FR-00009417.	
XX		
PA	26-JAN-2000; 2000FR-00001006.	
XX		
XX	(AVET ) AVENTIS CROPS SCIENCE SA.	
XX	(INRG ) INST NAT RECH AGRONOMIQUE.	

XX Beclin C, Elmayer T, Vaucheret H;  
 XX MPI; 2001-159529/16.  
 XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
 PT resistance in plants and, when inhibited, for increasing transgene  
 PT expression.  
 PS Claim 1; Page 31-32; 36pp; French.  
 XX The present sequence represents the genomic sequence of the Arabidopsis  
 CC thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional  
 CC inactivation (degradation of RNA) and for resistance to viruses.  
 CC Overexpression of SGS3 results in plants with increased resistance to  
 CC viruses, while inactivation of SGS3 in transgenic plants (e.g. by  
 CC expressing antisense RNA, by mutation or by homologous recombination)  
 CC increases the level of the transgene product. This product may e.g.  
 CC impart resistance (to herbicide, insects or pathogens), alter contents of  
 CC essential fatty acids or proteins, or is pharmaceutically active, e.g. an  
 CC immunoglobulin or interferon.  
 XX Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;  
 SQ  
 Query Match 100.0%; Score 3275; DB 4; Length 3275;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GACAAACAAACAAATTTAGCAATGTCATGTTGCTAGCAATTAATTAATGAGAAACA 60  
 1 GACAAACAAACAAATTTAGCAATGTCATGTTGCTAGCAATTAATTAATGAGAAACA 60  
 61 TTAAGTTAGCAAAAGGAAAAAGGATCAAAATCAAAATCAAACTGAAT 120  
 61 TTAAGTTAGCAAAAGGAAAAAGGATCAAAATCAAAATCAAACTGAAT 120  
 61 TTAAGTTAGCAAAAGGAAAAAGGATCAAAATCAAAATCAAACTGAAT 120  
 121 GAAAAATTGAGATCGAATCGAATAAAGAGCCGTTTGAAGCTTAATAGCTTCCTC 180  
 121 GAAAAATTGAGATCGAATCGAATAAAGAGCCGTTTGAAGCTTAATAGCTTCCTC 180  
 121 GAAAAATTGAGATCGAATCGAATAAAGAGCCGTTTGAAGCTTAATAGCTTCCTC 180  
 181 ATTGTCCTCTTCTGCTGATTAATTTCTTCTCGAGATCTGATCACTACTCTCAC 240  
 181 ATTGTCCTCTTCTGCTGATTAATTTCTTCTCGAGATCTGATCACTACTCTCAC 240  
 181 ATTGTCCTCTTCTGCTGATTAATTTCTTCTCGAGATCTGATCACTACTCTCAC 240  
 241 TCTCCGCGCTTAACTTAAGTCTCTGCTGTTTACTGTAAGTTTCTGCTTAAG 300  
 241 TCTCCGCGCTTAACTTAAGTCTCTGCTGTTTACTGTAAGTTTCTGCTTAAG 300  
 241 TCTCCGCGCTTAACTTAAGTCTCTGCTGTTTACTGTAAGTTTCTGCTTAAG 300  
 301 CTTCCGATGCGCTCAACGATGATCTGCTGATTTCTCTTTTCTTCTGCTGAAAA 360  
 301 CTTCCGATGCGCTCAACGATGATCTGCTGATTTCTCTTTTCTTCTGCTGAAAA 360  
 301 CTTCCGATGCGCTCAACGATGATCTGCTGATTTCTCTTTTCTTCTGCTGAAAA 360  
 361 ATTGCGCTTAATGTTCTGATTCGAAGTTTGTGTAGGTTACTTTTCTTCTTA 420  
 361 ATTGCGCTTAATGTTCTGATTCGAAGTTTGTGTAGGTTACTTTTCTTCTTA 420  
 361 ATTGCGCTTAATGTTCTGATTCGAAGTTTGTGTAGGTTACTTTTCTTCTTA 420  
 421 TTTTAATGTTCTTAAGTAAAGATACGATGCTGTAAGTTTGTGCTTCTTCTT 480  
 421 TTTTAATGTTCTTAAGTAAAGATACGATGCTGTAAGTTTGTGCTTCTTCTT 480  
 421 TTTTAATGTTCTTAAGTAAAGATACGATGCTGTAAGTTTGTGCTTCTTCTT 480  
 481 TTTACCGTTTATGCTGATCGAGATTTTGAATGTAAGTTTGTGCTTCTTCTT 540  
 481 TTTACCGTTTATGCTGATCGAGATTTTGAATGTAAGTTTGTGCTTCTTCTT 540  
 481 TTTACCGTTTATGCTGATCGAGATTTTGAATGTAAGTTTGTGCTTCTTCTT 540  
 541 TTTGTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 541 TTTGTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 541 TTTGTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 601 ATGCGTTGTTGCTGTTTAACTTCAAGTTCAATGTTGATGTTGATGTTGATG 660  
 601 ATGCGTTGTTGCTGTTTAACTTCAAGTTCAATGTTGATGTTGATGTTGATG 660  
 601 ATGCGTTGTTGCTGTTTAACTTCAAGTTCAATGTTGATGTTGATGTTGATG 660  
 661 ACTGTGGGTTTCTTGTGGCTAATGTTTAAATGATGTTCTAGGCTGATCAATGT 720

DB 661 ACTGTGGGTTTCTTGTGGCTAATGTTTAAATGATGTTCTAGGCTGATCAATGT 720  
 QY 721 CTAGGAAAGAAAGCTTCAGGTTGTTATAGGCTTGAAGTTTGAACGTTTCAAGTT 780  
 DB 721 CTAGGAAAGAAAGCTTCAGGTTGTTATAGGCTTGAAGTTTGAACGTTTCAAGTT 780  
 QY 781 TGGCAGGACGACGCTGCTTCTTCAAGATGATGAGAGAGAGTGAAGTCTTCA 840  
 DB 781 TGGCAGGACGACGCTGCTTCTTCAAGATGATGAGAGAGAGTGAAGTCTTCA 840  
 QY 841 AGAAGAAACAAACAAACGAAACCTTCTGAAAACTTGGGTTTCTCAATTCGA 900  
 DB 841 AGAAGAAACAAACAAACGAAACCTTCTGAAAACTTGGGTTTCTCAATTCGA 900  
 QY 901 ATCTCTTGAAGCTTGGGTTGTCAGCAGCAAGGAGAGTGAAGTCTTGGAGAG 960  
 DB 901 ATCTCTTGAAGCTTGGGTTGTCAGCAGCAAGGAGAGTGAAGTCTTGGAGAG 960  
 QY 961 GAAACAAATGATCCGAGAGAGTAAAGCAATGATGAGGAGTCAATCAATCTG 1020  
 DB 961 GAAACAAATGATCCGAGAGAGTAAAGCAATGATGAGGAGTCAATCAATCTG 1020  
 QY 1021 GTGGGAGCAGGCTTGAAGCAAGATGATGATGATGATGATGATGATGATGAT 1080  
 DB 1021 GTGGGAGCAGGCTTGAAGCAAGATGATGATGATGATGATGATGATGATGAT 1080  
 QY 1081 CTGGCCCTCTTGAAGAGAGATGAAATGAGCAAGAGAGAGTCTGTCAGCACA 1140  
 DB 1081 CTGGCCCTCTTGAAGAGAGATGAAATGAGCAAGAGAGAGTCTGTCAGCACA 1140  
 QY 1141 CAGCTGTCAGAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 DB 1141 CAGCTGTCAGAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 QY 1201 ATGATTCGATGCTTGAAGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
 DB 1201 ATGATTCGATGCTTGAAGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
 QY 1261 ATGTAGTCAAAAGAGCCATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
 DB 1261 ATGTAGTCAAAAGAGCCATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
 QY 1321 GCAAGTGGATGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
 DB 1321 GCAAGTGGATGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
 QY 1381 CAGCTTGTCAAGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 DB 1381 CAGCTTGTCAAGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
 QY 1441 ATGCGAGCAAAAGAGGCTAGGAGTTAAGCTCATGAGAAATGAGCTTAAGTTAG 1500  
 DB 1441 ATGCGAGCAAAAGAGGCTAGGAGTTAAGCTCATGAGAAATGAGCTTAAGTTAG 1500  
 QY 1501 AAAAGATCTAAGATGAGAGGCTCATCTTGTGAGATTTATGGCAGT 1560  
 DB 1501 AAAAGATCTAAGATGAGAGGCTCATCTTGTGAGATTTATGGCAGT 1560  
 QY 1561 GAAAGGTTGGTGAAGATGAAGAAAGATTAAGTTGTGAGCTCAATGTCATCA 1620  
 DB 1561 GAAAGGTTGGTGAAGATGAAGAAAGATTAAGTTGTGAGCTCAATGTCATCA 1620  
 QY 1621 TCAATTAATCTAGCTGATTAAGAGATTAAGATTAAGATTAAGATTAAGATTAAG 1680  
 DB 1621 TCAATTAATCTAGCTGATTAAGAGATTAAGATTAAGATTAAGATTAAGATTAAG 1680  
 QY 1681 TTTCTTAATTTCTCTGATCTAATCTAATCTAATCTAATCTAATCTAATCTAATCT 1740  
 DB 1681 TTTCTTAATTTCTCTGATCTAATCTAATCTAATCTAATCTAATCTAATCTAATCT 1740  
 QY 1741 ATGGCAACCAAGAGCTGGAATCTTGAACAAGTGAAGCTCTAGAGCAAGCCAT 1800  
 DB 1741 ATGGCAACCAAGAGCTGGAATCTTGAACAAGTGAAGCTCTAGAGCAAGCCAT 1800

```

Db 1741 ATGGGCAACCAAGCTGCTGGAAATCTTCGACAAATATAGAGCTCTTAGAGACCCCAT 1800
QY TCCATATGTCACAGGGCCATGCTGGAGATGAGTGTCTGATGTTTGAAGACAGTGCACAT 1860
Db 1801 TCCATATGTCACAGGGCCATGCTGGAGATGAGTGTCTGATGTTTGAAGACAGTGCACAT 1860
QY 1861 GGCATATTTGAGAGCCCAAGCCCTCCACCGGAGTTAGTGAATGGGCTTATAGTAAAT 1920
Db 1861 GGCATATTTGAGAGCCCAAGCCCTCCACCGGAGTTAGTGAATGGGCTTATAGTAAAT 1920
QY 1921 GCTGCGGGGTCAGAAAGCGCATGATGTTTCTGAGAGTGTTCGCCAATCTGATGGCTTCT 1980
Db 1921 GCTGCGGGGTCAGAAAGCGCATGATGTTTCTGAGAGTGTTCGCCAATCTGATGGCTTCT 1980
QY 1981 GCAACGAACCAAGATCTGACATATCAATCAACCTCAAGGCTCTCTCCCCCAAGA 2040
Db 1981 GCAACGAACCAAGATCTGACATATCAATCAACCTCAAGGCTCTCTCCCCCAAGA 2040
QY 2041 AATTTGATATATGCTTTTATGTTTGTCTATTTGAATTTAAAGTTTGTGGTCCGTGTA 2100
Db 2041 AATTTGATATATGCTTTTATGTTTGTCTATTTGAATTTAAAGTTTGTGGTCCGTGTA 2100
QY 2101 TGCATCTGTTATGATATATCTATGATTCATTAGGCAAAACAAAGCTGAATTCAGTTG 2160
Db 2101 TGCATCTGTTATGATATATCTATGATTCATTAGGCAAAACAAAGCTGAATTCAGTTG 2160
QY 2161 AATCATACCAAGATGCTGTAAGAGCTGAGGAGATCTCTAGGACAAATCAGCAG 2220
Db 2161 AATCATACCAAGATGCTGTAAGAGCTGAGGAGATCTCTAGGACAAATCAGCAG 2220
QY 2221 CTGAACCTATTTAAGAACAGCTCTCAAAACAGAACAGACCCCAAGGCTTTGAGAA 2280
Db 2221 CTGAACCTATTTAAGAACAGCTCTCAAAACAGAACAGACCCCAAGGCTTTGAGAA 2280
QY 2281 TCTCTGGAATATATGAGCGAAGAGCTGCTGACAACTGCAAGGATATCGATCGTGA 2340
Db 2281 TCTCTGGAATATATGAGCGAAGAGCTGCTGACAACTGCAAGGATATCGATCGTGA 2340
QY 2341 CAGAGAACTAAGATGAGCATGAACGAAACAGGAGAGATGATTTTCTAGAAAT 2400
Db 2341 CAGAGAACTAAGATGAGCATGAACGAAACAGGAGAGATGATTTTCTAGAAAT 2400
QY 2401 CACAAACTTGACATTTTGTATTAACCTACTGATTCACATTTTGTATATATGTCACAA 2460
Db 2401 CACAAACTTGACATTTTGTATTAACCTACTGATTCACATTTTGTATATATGTCACAA 2460
QY 2461 AAAACCTGTGTGTTTGAAGATGATGACACAGAGGTTTTCATGATTCATCAAA 2520
Db 2461 AAAACCTGTGTGTTTGAAGATGATGACACAGAGGTTTTCATGATTCATCAAA 2520
QY 2521 CAGATCCAGAAAGAAAGAACGCAAGGAGGAAATTTGAGATGTTTGAAGCAGAGAA 2580
Db 2521 CAGATCCAGAAAGAAAGAACGCAAGGAGGAAATTTGAGATGTTTGAAGCAGAGAA 2580
QY 2581 CGTGCCAAGGTTTGTGCGCAGACGACGAGCAATTAATCCCTCTAGCAATGCAATTC 2640
Db 2581 CGTGCCAAGGTTTGTGCGCAGACGACGAGCAATTAATCCCTCTAGCAATGCAATTC 2640
QY 2641 CGAAAGAGATATATGATCTAATCAATATCCCTCTGCGGTTTGTGTTTCAAACTA 2700
Db 2641 CGAAAGAGATATATGATCTAATCAATATCCCTCTGCGGTTTGTGTTTCAAACTA 2700
QY 2701 AAGATTAAGATATATCCGGTGTGATCTTCTGAGAGCTGAGGAAGTCAACCTTC 2760
Db 2701 AAGATTAAGATATATCCGGTGTGATCTTCTGAGAGCTGAGGAAGTCAACCTTC 2760
QY 2761 ATCGAGTTTCAAGAGAAAGATGAGAGTGTGAGAGAGGAGAGATGCTGATATAA 2820
Db 2761 ATCGAGTTTCAAGAGAAAGATGAGAGTGTGAGAGAGGAGAGATGCTGATATAA 2820
QY 2821 GATCAAGAGAAAGATGAGAGATGAGAGATGAGAGATGAGAGATGATGATCTG 2880
Db 2821 GATCAAGAGAAAGATGAGAGATGAGAGATGAGAGATGAGAGATGATGATCTG 2880

```

```

QY 2881 GAGAAAGATTTGATGAGGCTTTGGAAACAGCTCATGTACAAACATGAGCTTCAATGAA 2940
Db 2881 GAGAAAGATTTGATGAGGCTTTGGAAACAGCTCATGTACAAACATGAGCTTCAATGAA 2940
QY 2941 GATGATTGAGACAAAGTCTGTATCAACAGACAAAGTCTTGTGTTTGTCTTTT 3000
Db 2941 GATGATTGAGACAAAGTCTGTATCAACAGACAAAGTCTTGTGTTTGTCTTTT 3000
QY 3001 GATGTCGGAAGTATGAGAAATCTGAGAGCTCATTAAATCTAGGACAAATCTAAGA 3060
Db 3001 GATGTCGGAAGTATGAGAAATCTGAGAGCTCATTAAATCTAGGACAAATCTAAGA 3060
QY 3061 GATTATAGATTTATATCTCCCAATTTTGTAGACGATCTAAGAGATTAAGTTCTT 3120
Db 3061 GATTATAGATTTATATCTCCCAATTTTGTAGACGATCTAAGAGATTAAGTTCTT 3120
QY 3121 GTGACTAAACCAAGTTCTTATGATATTTTGTGTTTGTGTTTGTGTTTGTGTTT 3180
Db 3121 GTGACTAAACCAAGTTCTTATGATATTTTGTGTTTGTGTTTGTGTTTGTGTTT 3180
QY 3181 TTAGACATTTTACCAAGCTCAGAGTGAATCAAGAAATGGAATCAATGATGTTT 3240
Db 3181 TTAGACATTTTACCAAGCTCAGAGTGAATCAAGAAATGGAATCAATGATGTTT 3240
QY 3241 AGAATTTTATCTACAAATATATATGCTAACAAT 3275
Db 3241 AGAATTTTATCTACAAATATATATGCTAACAAT 3275

RESULT 2
AAF25374
ID AAF25374 standard; cDNA; 1878 BP.
XX
AC AAF25374;
XX
DT 15-MAY-2001 (first entry)
XX
DE Nucleotide sequence of the Arabidopsis SGS3 polypeptide.
XX
KW SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX viral resistance; resistance; fatty acid content; protein content; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 1..1878
FT /tag= a
FT /product= "SGS3"
XX
PN MO200105951-A2.
XX
PD 25-JAN-2001.
XX
PF 13-JUL-2000; 2000WO-FR002052.
XX
PR 16-JUL-1999; 99FR-00009417.
XX
PR 26-JAN-2000; 2000FR-00001006.
XX
PA (AVET ) AVENTIS CROPS/SCIENCE SA.
XX (INRG ) INST NAT RECH AGRONOMICUE.
XX
PI Beclin C, Elmayer T, Vaucheret H;
XX
DR WPI; 2001-159529/16.
XX P-PSDB; AAB31798.
XX
PT New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
PT resistance in plants and, when inhibited, for increasing transgene
PT expression.
XX
PS Claim 1; Page 32-35; 36pp; French.
XX

```

The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide. The SGS3 gene is essential for post-transcriptional inactivation of (degradation of RNA) and for resistance to viruses. Overexpression of SGS3 results in plants with increased resistance to viruses, while inactivation of SGS3 in transgenic plants (e.g. by expressing antisense RNA, by mutation or by homologous recombination) increases the level of the transgene product. This product may e.g. impart resistance to herbicide, insects or pathogens, alter contents of essential fatty acids or proteins, or is pharmaceutically active, e.g. an immunoglobulin or interferon

Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;

Query Match 29.4%; Score 963; DB 4; Length 1878;  
Best Local Similarity 100.0%; Pred. No. 0; Indels 0; Gaps 0;  
Matches 963; Conservative 0; Mismatches 0

```

Oy 696 ATGAGTTCTAGGCTGTGCTCAATGCTTAAGAAAGAAAGCTTCAGGGTGTATAGGCT 755
Db 1 ATGAGTTCTAGGCTGTGCTCAATGCTTAAGAAAGAAAGCTTCAGGGTGTATAGGCT 60
Oy 756 GAGGTGAACAGTGTGTTCAAGTTTGGCAGGAGCACTGGCTTCTTCAAGATGAT 815
Db 61 GAGGTGAACAGTGTGTTCAAGTTTGGCAGGAGCACTGGCTTCTTCAAGATGAT 120
Oy 816 GAGGAGAGTGGAGTGTCTTCCAAAGAAAGAAAGCAACAGAAACCTTCTGGA 875
Db 121 GAGGAGAGTGGAGTGTCTTCCAAAGAAAGAAAGCAACAGAAACCTTCTGGA 180
Oy 876 AAACTTGGGTTTCTCAGAAATTCCTCTAGAGCTTGGGTGTCTCAGCAGCAAGG 935
Db 181 AAACTTGGGTTTCTCAGAAATTCCTCTAGAGCTTGGGTGTCTCAGCAGCAAGG 240
Oy 936 AGAGTAGCAACGTAATCTGGGAGAGAAACAATATCCGGAGAGGTAAACGCAATGT 995
Db 241 AGAGTAGCAACGTAATCTGGGAGAGAAACAATATCCGGAGAGGTAAACGCAATGT 300
Oy 996 CGGGGATTCAGAGTAACATATCTGTGCGGAGCAGCGCTTGAGCAAGAAATGATAC 1055
Db 301 CGGGGATTCAGAGTAACATATCTGTGCGGAGCAGCGCTTGAGCAAGAAATGATAC 360
Oy 1056 AACTTTGTGGCAACCCCACTGTATCTGCGCTCTTTGGAAGAGATGATGGCAG 1115
Db 361 AACTTTGTGGCAACCCCACTGTATCTGCGCTCTTTGGAAGAGATGATGGCAG 420
Oy 1116 GCAAGAGAGAGTCTGCTCAGCAACAGCTGTGAGAGATTTCTGACGTGAGATAT 1175
Db 421 GCAAGAGAGAGTCTGCTCAGCAACAGCTGTGAGAGATTTCTGACGTGAGATAT 480
Oy 1176 GTGATATATCTCTGAGAGAGATGATTCGAGCTTGGATGATCTGATGACGAC 1235
Db 481 GTGATATATCTCTGAGAGAGATGATTCGAGCTTGGATGATCTGATGACGAC 540
Oy 1236 CTTCAGATGATATGATGATCTGAGATGTGATCAAAAGCAGATGATCAAGAGCAG 1295
Db 541 CTTCAGATGATATGATGATCTGAGATGTGATCAAAAGCAGATGATCAAGAGCAG 600
Oy 1296 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGGATGATGATGATGATGATGAT 1355
Db 601 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGGATGATGATGATGATGATGAT 660
Oy 1356 GAACCAAGAGAGAGTGGCATTTGTCAGCTTGTGAGAGCAGACTGTGTCATGATGG 1415
Db 661 GAACCAAGAGAGAGTGGCATTTGTCAGCTTGTGAGAGCAGACTGTGTCATGATGG 720
Oy 1416 TATAACCTGCACCTCTAAGCTCATGCGAGCAAAAGAGAGTAGCGAGTTAAGCTC 1475
Db 721 TATAACCTGCACCTCTAAGCTCATGCGAGCAAAAGAGAGTAGCGAGTTAAGCTC 780
Oy 1476 CATAGAGAAATGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGCGCATCTGAT 1535
Db 781 CATAGAGAAATGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGCGCATCTGAT 840

```

```

Oy 1536 CTTGTGTGATGATTTATGGGCACTGAGAGGGTTTGGTGGAGATGAAAGATTAATGA 1595
Db 841 CTTGTGTGATGATTTATGGGCACTGAGAGGGTTTGGTGGAGATGAAAGATTAATGA 900
Oy 1596 ATTGTGCGCTCCATGATGATCATGATGATGATGATGATGATGATGATGATGAT 1655
Db 901 ATTGTGCGCTCCATGATGATCATGATGATGATGATGATGATGATGATGATGAT 960
Oy 1656 AAG 1658
Db 961 AAG 963

```

RESULT 3  
AAF25372/c  
ID AAF25372 standard, DNA, 27 BP.

AC AAF25372;  
XX 15-MAY-2001 (first entry)  
XX PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.  
XX SGS3 gene; post-transcriptional inactivation; RNA degradation;  
XX viral resistance; resistance; fatty acid content; protein content;  
XX PCR primer; ss.

XX Arabidopsis thaliana.

XX WO200105951-A2.

XX 25-JAN-2001.

XX 13-JUL-2000; 2000WO-FR002052.

XX 16-JUL-1999; 99FR-00009417.

XX 26-JAN-2000; 2000FR-00001006.

XX (AVET ) AVENTIS CROPS/SCIENCE SA.  
XX (INRG ) INST NAT RECH AGRONOMIQUE.

XX Beclin C, Elmayan T, Vaucheret H;

XX WPI; 2001-159529/16.

XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
XX resistance in plants and, when inhibited, for increasing transgene  
XX expression.

XX Example 1; Page 22; 36pp; French.

XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis  
XX thaliana SGS3 polypeptide. The SGS3 gene is essential for post-  
XX transcriptional inactivation (degradation of RNA) and for resistance to  
XX viruses. Overexpression of SGS3 results in plants with increased  
XX resistance to viruses, while inactivation of SGS3 in transgenic plants  
XX (e.g. by expressing antisense RNA, by mutation or by homologous  
XX recombination) increases the level of the transgene product. This product  
XX may e.g. impart resistance (to herbicide, insects or pathogens), alter  
XX contents of essential fatty acids or proteins, or is pharmaceutically  
XX active, e.g. an immunoglobulin or interferon

XX Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;

Query Match 0.8%; Score 27; DB 4; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.055; Indels 0; Gaps 0;  
Matches 27; Conservative 0; Mismatches 0

```

Oy 2926 GGCCTTCAATGAAGATGATGAGAC 2952
Db 27 GGCCTTCAATGAAGATGATGAGAC 1

```



```

RESULT 4
AAF25371
XX AAF25371 standard; DNA; 23 BP.
AC AAF25371;
XX
XX 15-MAY-2001 (first entry)
XX
XX PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
DE
XX
XX SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX viral resistance; resistance; fatty acid content; protein content;
XX PCR primer; ss.
XX
XX Arabidopsis thaliana.
XX
XX WO200105951-A2.
XX
XX 25-JAN-2001.
XX
XX 13-JUL-2000; 2000MO-FR002052.
XX
XX 16-JUL-1999; 99FR-00009417.
XX 26-JAN-2000; 2000FR-00001006.
XX
XX (AVERTIS ) AVENTIS CROSCIENCE SA.
XX (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX Becin C, Elmayan T, Vaucheret H;
XX
XX WPI; 2001-159529/16.
XX
XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
XX resistance in plants and, when inhibited, for increasing transgene
XX expression.
XX
XX Example 1; Page 22; 36pp; French.
XX
XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
XX thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
XX transcriptional inactivation (degradation of RNA) and for resistance to
XX viruses. Overexpression of SGS3 results in plants with increased
XX resistance to viruses, while inactivation of SGS3 in transgenic plants
XX (e.g. by expressing antisense RNA, by mutation or by homologous
XX recombination) increases the level of the transgene product. This product
XX may e.g. impart resistance (to herbicide, insects or pathogens), alter
XX contents of essential fatty acids or proteins, or is pharmaceutically
XX active, e.g. an immunoglobulin or interferon
XX
XX
XX Sequence 23 BP; 6 A; 4 C; 7 G; 6 T; 0 U; 0 Other;
SQ
Query Match 0.7%; Score 23; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 693 AAAATGAGTTCTAGGCGTGTCC 715
DB 1 AAAATGAGTTCTAGGCGTGTCC 23

```

```

XX
XX Homo sapiens.
XX
XX WO200202806-A2.
XX
XX 10-JAN-2002.
XX
XX 29-JUN-2001; 2001WO-EP007470.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piegenbrock C, Berlin K;
XX
XX WPI; 2002-154757/20.
XX
XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
XX useful for detecting cytosine methylation state of genes associated with
XX pharmacogenomics and for therapy of diseases e.g. cancer.
XX
XX Claim 1; SEQ ID NO 123; 24pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence at least 18
XX bases in length of a segment of the chemically pretreated DNA of genes
XX associated with pharmacogenomics according to one of the sequences of the
XX genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
XX (NM_000776 and NM_017460), DYPD (NM_000110), EPHX2 (NM_001979), OCLN
XX (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_004996), OCLN
XX NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899 and
XX their complementary sequences, or a sequence (S1) chosen from 87
XX sequences and their complements. The chemical pretreatment is bisulphite
XX treatment to convert cytosines (but not methyl-cytosines) into uracils.
XX Also included are an oligomer (II) in particular an oligonucleotide or a
XX peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
XX base sequence having a length of 9 nucleotides which hybridises to or is
XX identical to a chemically pretreated DNA of genes associated with
XX pharmacogenomics and their complements, arranged in an array for
XX analysing diseases associated with the methylation state (CpG) and/or
XX detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
XX oligomers may also be used as PCR primers. The set of 87 nucleic acids
XX and their complements is useful for diagnosis and therapy of solid
XX tumours and cancer. The present sequence represents one the 87 DNA
XX sequences or its complement. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 6136 BP; 1661 A; 69 C; 1415 G; 2991 T; 0 U; 0 Other;
SQ
Query Match 0.7%; Score 22; DB 6; Length 6136;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3145 TATTTGTTTGTGTA 3166
DB 2122 TATTTGTTTGTGTA 2143

```

```

RESULT 5
ABK40041
ID ABK40041 standard; DNA; 6136 BP.
XX
XX ABK40041;
XX
XX 21-MAY-2002 (first entry)
XX
XX Human chemically pretreated gene sequence #62 strand 1.
XX
XX Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
XX cytostatic; ALDH6; CYP11A; CYP11B1; DYPD; EPHX2; OCLN; TXNRD1;
XX UGT8; WBP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX

```

```

RESULT 6
ABN80100
ID ABN80100 standard; DNA; 6301 BP.
XX
XX ABN80100;
XX
XX 15-JUL-2002 (first entry)
XX
XX Human chemically modified disease associated gene SEQ ID NO 117.
XX
XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
XX heart disease; epilepsy; histone deacetylation; muscular dystrophy;
XX dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
XX

```



```

XX 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX (EPIC-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
PI WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX Claim 1; SEQ ID NO 482; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/vulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6048 BP; 1296 A; 103 C; 1361 G; 3288 T; 0 U; 0 Other;
Query Match 0.6%; Score 21; DB 6; Length 6048;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 526 TTGGTTTGGTTTGGTTT 546
Db 5395 TTGGTTTGGTTTGGTTT 5415
RESULT 9
ABL32031
ID ABL32031 standard; DNA; 6072 BP.
XX ABL32031;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Human immune system associated gene SEQ ID NO: 4.
DE
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemic; cytosatic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
XX Homo sapiens.
OS
XX WO200200928-A2.
PN
XX
XX 03-JAN-2002.
PD
XX
XX 02-JUL-2001; 2001WO-EP007537.
PF
XX 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX (EPIC-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful for
PT

```

```

PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX Claim 1; SEQ ID NO 4; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/vulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6072 BP; 1595 A; 155 C; 1384 G; 2938 T; 0 U; 0 Other;
Query Match 0.6%; Score 21; DB 6; Length 6072;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3145 TATTTTGGTTTGGTTTGGTTA 3165
Db 3717 TATTTTGGTTTGGTTTGGTTA 3737
RESULT 10
AAS46448
ID AAS46448 standard; DNA; 8245 BP.
XX AAS46448;
AC
XX
XX 18-DEC-2001 (first entry)
DT
XX
XX Tumour suppressor gene derived chemically modified sequence #170.
DE
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytosatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; de.
XX
XX Homo sapiens.
OS
XX WO200168912-A2.
PN
XX
XX 20-SEP-2001.
PD
XX
XX 15-MAR-2001; 2001WO-EP002955.
PF
XX 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX (EPIC-) EPIGENOMICS AG.
PA
XX Olek A, Piepenbrock C, Berlin K;
PI WPI; 2001-602752/68.
XX Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX Claim 1; SEQ ID NO 170; 27pp; English.
XX The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single

```

CC nucleotide polymorphisms and also to be used in an array for analysing  
 CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The  
 CC probes can also be used in a method for ascertaining genetic and/or  
 CC epigenetic parameters for the diagnosis and/or therapy of existing  
 CC diseases or the predisposition to specific diseases, by analysing  
 CC cytosine methylations. The parameters may be compared to another set of  
 CC genetic and/or epigenetic parameters, the differences serving as basis  
 CC for diagnosis and/or prognosis events which are disadvantageous to  
 CC patients. The present sequence is one of the 533 genomic sequences  
 CC derived from tumour suppressor genes and oncogenes. Sequences with even  
 CC numbered Seq ID numbers are the complementary sequence of the  
 CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID  
 CC 535, except for those whose partner sequence is missing). Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIGO  
 CC at [http://wigo.int/pub/published\\_pct\\_sequences](http://wigo.int/pub/published_pct_sequences)

XX Sequence 8245 BP; 1978 A; 305 C; 2119 G; 3842 T; 0 U; 1 Other;

Query Match 0.6%; Score 21; DB 4; Length 8245;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3141 TTAGTATTGTTTTTTTGG 3161

Db 6632 TTAGTATTGTTTTTTTGG 6652

#### RESULT 11

ABN80176 standard; DNA; 13453 BP.

AC ABN80176;

XX 15-JUL-2002 (first entry)

DE Human chemically modified disease associated gene SEQ ID NO 193.

XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;

KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;

KM dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;

XX antidiabetic; cytosine; anticonvulsant; de.

OS Homo sapiens.

XX Synthetic.

XX WO200200927-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007536.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130908/17.

XX Novel nucleic acid useful for diagnosis and therapy of diseases

XX associated with development genes such as diabetes, comprises a sequence

XX of a segment of chemically pretreated DNA of genes associated with

XX development.

XX Claim 1; SEQ ID NO 193; 27bp; English.

XX The invention relates to a nucleic acid (I) comprising a sequence at

XX least 18 bases in length of a segment of chemically pretreated DNA (II)

XX of genes associated with development selected from 87 genes listed in the

XX specification such as ACCPN, ADFN, or APD1 and comprising one of 350

XX sequences (ABN79984-ABN80333) or their complements. The invention is

CC useful for the diagnosis or therapy of diseases associated with  
 CC development genes, in particular disease related to homeobox containing  
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
 CC associated with congenital heart disease, epilepsy, diseases related to  
 CC histone deacetylation, Currarino syndrome, diseases related with the  
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.

CC Oligomers specific to each of the genes are useful for detecting the

CC methylation state of all CpG dinucleotides within the 350 sequences or

CC (II) and their complementary sequences, (II) and/or their complements and

CC the amplification of the 350 sequences, (II) and/or their complements and

CC as oligomer probes for detecting the cytosine methylation state and/or

CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this

CC patent did not form part of the printed specification but is based on

CC sequence information supplied to Derwent by the European Patent Office

XX Sequence 13453 BP; 2569 A; 384 C; 3712 G; 6788 T; 0 U; 0 Other;

Query Match 0.6%; Score 21; DB 6; Length 13453;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3141 TTAGTATTGTTTTTTTGG 3161

Db 7909 TTAGTATTGTTTTTTTGG 7929

#### RESULT 12

ADA02738/C standard; DNA; 52302 BP.

AC ADA02738;

XX 06-NOV-2003 (first entry)

DE Human CCND2 carcinoma associated gene, SEQ ID NO:1256.

XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;

KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;

XX gene; ds.

OS Homo sapiens.

XX WO2003057146-A2.

XX 17-JUL-2003.

XX 26-DEC-2002; 2002WO-US041414.

XX 26-DEC-2001; 2001US-00035832.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-587068/55.

XX New recombinant nucleic acid encoding carcinoma associated protein,

XX useful for preparing compositions for treating carcinomas.

XX Claim 1; SEQ ID NO 1256; 245bp; English.

XX The invention relates to recombinant carcinoma associated (CA) nucleic

XX acid sequences from mouse and human (ADA01482-ADA03094), and to

XX recombinant carcinoma associated proteins (CAP) encoded by them. The

XX invention also encompasses expression vectors and host cells comprising a

XX CA nucleic acid, a polypeptide (especially an antibody) that specifically

XX binds to the protein, and a biochip comprising CA nucleic acid or

XX fragments thereof. The sequences of the invention were identified using

XX oncogenic retroviruses, which insert into the genome of the host organism

XX at random. Many of these do not carry transduced host oncogenes or

XX pathogenic trans-acting viral genes, meaning that cancer incidence is a

XX direct consequence of the effects of proviral integration into host

XX protooncogenes. The CA nucleic acid sequences can be used to diagnose

CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed human CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 8; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2981 GTTCTTTGTTTGCTTTGG 3001

Db 11586 GTTCTTTGTTTGCTTTGG 11566

RESULT 13

ADB72476/c

ID ADB72476 standard; DNA; 52302 BP.

XX ADB72476;

AC 04-DEC-2003 (first entry)

XX Human CCND2 gene.

DE Human CCND2 gene.

XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;

KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX Homo sapiens.

OS Homo sapiens.

XX WO2003008583-A2.

PN 30-JAN-2003.

XX 26-DEC-2001; 2001WO-US051291.

PF 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-00004113.

PR 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00997722.

PR 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

PA Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

DR New recombinant nucleic acid, useful for treating carcinomas, lymphomas,

XX cancers, neoplasm, adenocarcinoma, or sarcomas.

PT Claim 1; SEQ ID NO 304; 2304pp; English.

PS The invention relates to a novel recombinant nucleic acid comprising a

XX nucleotide sequence selected from any of the 660 sequences fully defined

CC in the specification. A polynucleotide of the invention has cytostatic

CC activity, and may have a use in gene therapy, or in a vaccine. The

CC recombinant nucleic acids and polypeptides are useful for treating

CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and

CC sarcomas. The present sequence represents a human gene of the invention.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 9; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2981 GTTCTTTGTTTGCTTTGG 3001

Db 11586 GTTCTTTGTTTGCTTTGG 11566

RESULT 14

ADC85218/c

ID ADC85218 standard; DNA; 52302 BP.

XX ADC85218;

AC 01-JAN-2004 (first entry)

XX Human CCnd2 genomic sequence.

DE Human CCnd2 genomic sequence.

XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;

KW secreted; transmembrane; intracellular; ds.

XX Homo sapiens.

OS Homo sapiens.

XX WO2003045230-A2.

PN 05-JUN-2003.

XX 02-DEC-2002; 2002WO-US038582.

PF 30-NOV-2001; 2001US-00997722.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-513603/48.

DR The invention relates to a recombinant nucleic acid comprising a

XX nucleotide sequence selected from any of the fully defined carcinoma-

CC associated (CA) genes from the 50 tables given in the specification. The

CC CA proteins are secreted, transmembrane or intracellular proteins. The

CC recombinant nucleic acids are useful for screening for drug candidates

CC for diagnosing or treating carcinomas. Sequences given in ADC85215-

CC ADC85514 represent CA genes of the invention.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 9; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2981 GTTCTTTGTTTGCTTTGG 3001

Db 11586 GTTCTTTGTTTGCTTTGG 11566

RESULT 15

AAC52596/c

ID AAC52596 standard; DNA; 420 BP.

XX AAC52596;

AC 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 71849.

DE Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway; metabolic pathway;

KW promoter; termination sequence; ss.



PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154779P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 23-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155659P.  
 PR 28-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-0156596P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161820P.  
 PR 28-OCT-1999; 99US-0161922P.  
 PR 28-OCT-1999; 99US-0161933P.  
 PR 29-OCT-1999; 99US-0162142P.

Query Match 0.6%; Score 20; DB 3; Length 420;  
 Best Local Similarity 100.0%; Pred.No. 1.1e+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACACACACAAATTTAG 21  
 |||||  
 Db 378 ACACACACAAATTTAG 359

Search completed: April 9, 2004, 02:09:03  
 Job time : 842.022 secs



**This Page Blank (uspto)**





AU236368  
 LOCUS AU236368 647 bp mRNA linear EST 01-APR-2002  
 DEFINITION AU236368 RAF1.4 Arabidopsis thaliana cDNA clone RAF1.4-93-K05 5', mRNA sequence.  
 ACCESSION AU236368  
 VERSION AU236368.1 GI:19875537  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 647)  
 Seki, M., Nakuraka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Oono, Y., Saito, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.  
 Large scale analysis of Arabidopsis full-length cDNA (2002b)  
 Unpublished (2002)  
 CONTACT: Motoaki Seki  
 Plant Functional Genomics Research Group  
 RIKEN Genomic Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-4359  
 Fax: 81-298-36-9060  
 Email: mseki@rc.riken.go.jp  
 An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FIC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified plasmid vector. Please visit our web site ([http://www.gsc.riken.go.jp/e/plan/index\\_e.html](http://www.gsc.riken.go.jp/e/plan/index_e.html)) for further details.

## FEATURES

source

Location/Qualifiers  
 1..647  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3702"  
 /clone="RAF1.4-93-K05"  
 /isue\_type="root"  
 /lab\_host="DH10B"  
 /clone\_lib="RAF1.4"  
 /note="Site\_1: BamHI; Site\_2: SalI"

## ORIGIN

Query Match 15.9%; Score 521; DB 9; Length 647;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-249;  
 Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

687 GTTGTAAATGAGTCTAGAGGCTGCTCAATGCTTAGAGAAAGAAACGTTCAAGGGTGGT 746  
 83 GTTGTAAATGAGTCTAGAGGCTGCTCAATGCTTAGAGAAAGAAACGTTCAAGGGTGGT 142  
 747 TATAGGCTGAGGTTGAACAGTTGTTCAAGGTTTGGCAGGGAAGAGTGGTCTTCA 806  
 143 TATAGGCTGAGGTTGAACAGTTGTTCAAGGTTTGGCAGGGAAGAGTGGTCTTCA 202  
 807 CAAAGATGAGGAGAGTGGAGTCAATTTCCAAAGAAACAAAGAACCAAGAAC 866  
 203 CAAAGATGAGGAGAGTGGAGTCAATTTCCAAAGAAACAAAGAACCAAGAAC 262  
 867 ACTTTCGAAAACTGGGTTTCTCAGAAATCGAATCTCTAGAGCTTGGGGTGTGAG 926  
 263 ACTTTCGAAAACTGGGTTTCTCAGAAATCGAATCTCTAGAGCTTGGGGTGTGAG 322  
 927 CAGCAAGGAGAGTGAACATGATCTGGAGAGAGAAACATGATCTGGAGAGTAAAC 986  
 323 CAGCAAGGAGAGTGAACATGATCTGGAGAGAGAGAAACATGATCTGGAGAGTAAAC 382  
 987 GGCAATGCTGGGGGCAATTAAGCTAACATATCTGGTTCGGGAGAGAGCTTTGAGCAAAAG 1046  
 383 GGCAATGCTGGGGGCAATTAAGCTAACATATCTGGTTCGGGAGAGAGCTTTGAGCAAAAG 442

OY 1047 TATGATTAACACTTGTGGACCCCACTGATATCTCGCCTCTTTGGAGAGAGATGG 1106  
 DB 443 TATGATTAACACTTGTGGACCCCACTGATATCTCGCCTCTTTGGAGAGAGATGG 502  
 OY 1107 AATTGGCAGCAAGAGAGGTTCTGCTCAGCAACACAGCTGTCAGAGATTTCTGACGTG 1166  
 DB 503 AATTGGCAGCAAGAGAGGTTCTGCTCAGCAACACAGCTGTCAGAGATTTCTGACGTG 562  
 OY 1167 GAGGATGATGATTAATGCTTCTGAGAGAGAGATGATTC 1207  
 DB 563 GAGGATGATGATTAATGCTTCTGAGAGAGAGATGATTC 603

RESULT 4  
 LOCUS AO964580/c 613 bp DNA linear GSS 28-JUN-2000  
 DEFINITION LERX20TF LERG Arabidopsis thaliana genomic clone LERX20, genomic survey sequence.  
 ACCESSION AO964580  
 VERSION AO964580.1 GI:6792281  
 KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

## ORGANISM

Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 613)

## AUTHORS

Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utterbach, T., Feilblyum, T., Liang, P., Creasy, T., and Fraser, C.M.  
 Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms  
 Unpublished (2000)

## JOURNAL

COMMENT  
 CONTACT: Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: atetigr.org  
 For additional information, see <http://www.tigr.org/tdb/ac/at.html>  
 Seq primer: 7P  
 Class: Shotgun.

FEATURES  
 source Location/Qualifiers

1..613  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="LANDSBERG ERRECTA"  
 /db\_xref="taxon:3702"  
 /clone="LERX20"  
 /clone\_lib="LERG"  
 /note="Torgan: Leaf; Vector: pUC19UR; Total genomic DNA was sheared to 0.4-0.7 Kbp before ligation."

## ORIGIN

Query Match 15.4%; Score 504; DB 28; Length 613;  
 Best Local Similarity 99.7%; Pred. No. 3.8e-241;  
 Matches 604; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

615 TGTATTAACCTTCAAGCTTCAATGTTGATTTGAGATTTTGTAGTCACTGGGTTTCTT 674  
 606 TGTATTAACCTTCAAGCTTCAATGTTGATTTGAGATTTTGTAGTCACTGGGTTTCTT 547  
 675 TGTGAGTATAGTGTAT 734  
 546 TGTGAGTATAGTGTAT 487  
 735 GTTCAGGAGTGTATATAGGCTGAGGTTGAACAGTTGTTCAAGTTTGGCAGGAGCAGAG 794  
 486 GTTCAGGAGTGTATATAGGCTGAGGTTGAACAGTTGTTCAAGTTTGGCAGGAGCAGAG 427  
 795 CTGGCTTTCTTCAAGATGATGAGAGAGTGGAGGTCATTTCCAGAAAGAACCAAGAAC 854  
 426 CTGGCTTTCTTCAAGATGATGAGAGAGTGGAGGTCATTTCCAGAAAGAACCAAGAAC 367

QY 855 AAACGAGGAAACACTTCTGGAAGAACTTGGGTTTCTCAGAAATTCGAATCTCTAGAGCT 914  
 Db 366 AAACGAGGAAACACTTCTGGAAGAACTTGGGTTTCTCAGAAATTCGAATCTCTAGAGCT 307  
 QY 915 TGGGGTGGTCCAGCAGCAGGAGAGGTAGCAAGTATCTGGAGAGAGAAACAATGATATCC 974  
 Db 306 TGGGGTGGTCCAGCAGCAGGAGAGGTAGCAAGTATCTGGAGAGAGAAACAATGATATCC 247  
 QY 975 GGGAGAGGTAAACGAGCAATGGTGGGGCATTTCAAGCTAAACATATCTGGTGGGAGAGAGCG 1034  
 Db 246 GGGAGAGGTAAACGAGCAATGGTGGGGCATTTCAAGCTAAACATATCTGGTGGGAGAGAGCG 187  
 QY 1035 TTGAGCAGAAATATATATATTAACAATTTTGGGACACCCCACTGTATCTCCGCTCTTGG 1094  
 Db 186 TTGAGCAGAAATATATATATTAACAATTTTGGGACACCCCACTGTATCTCCGCTCTTGG 127  
 QY 1095 GAAGAGAGATGGAATTTGAGGAGGAGAGAGAGGTTTCTGCTCAGCAGCAGAGCTGTGAGAG 1154  
 Db 126 GAAGAGAGATGGAATTTGAGGAGGAGAGAGAGGTTTCTGCTCAGCAGCAGAGCTGTGAGAG 67  
 QY 1155 TTTCCTGACGTGAGAGATGATGATTAATGCTTCTGAGAGAGAGATGATTCGATGCT 1214  
 Db 66 TTTCCTGACGTGAGAGATGATGATTAATGCTTCTGAGAGAGAGATGATTCGATGCT 7  
 QY 1215 TTGGAT 1220  
 Db 6 TTGGAT 1

RESULT 5  
 CENS00P9H 457 bp DNA linear GSS 28-JUN-1999  
 LOCUS Arabidopsis thaliana genome survey sequence T7 end of BAC F821 of  
 DEFINITION IGF library from strain Columbia of Arabidopsis thaliana, genomic  
 survey sequence.  
 AL084227  
 AL084227.1 GI:5285367

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 457)  
 Samsom,D., Saurin,M., Choisme,N., Artiguenave,F., Brothier,P., Wincker,P.,  
 JOURNAL  
 AUTHORS  
 REFERENCE  
 JOURNAL  
 TITLES  
 JOURNAL  
 Direct Submission  
 Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 Location/Qualifiers  
 1..457  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="F821"  
 /clone\_1b="IGF"  
 /note="end : T7"

ORIGIN  
 Query Match 13.8%; Score 451; DB 29; Length 457;  
 Best local Similarity 100.0%; Pred. No. 1.6e-214;  
 Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1682 TCTTTAATTTTCTCTGCAATCTACTGATCTTAGAATGTTAATGTTAGTCTGGCA 1741  
 Db 7 TCTTTAATTTTCTCTGCAATCTACTGATCTTAGAATGTTAATGTTAGTCTGGCA 66

QY 1742 TGGCAACCAAGAGCTGCTGGAATATCTTCAGCAAGTATGAGGCTTTAGAGACGCCATT 1801  
 Db 67 TGGCAACCAAGAGCTGCTGGAATATCTTCAGCAAGTATGAGGCTTTAGAGACGCCATT 126  
 QY 1802 CCTATGTCACAGGCGCCATGCTGGATGATGTTTCTGATGTTTGAAGACAGTGGCACTG 1861  
 Db 127 CCTATGTCACAGGCGCCATGCTGGATGATGTTTCTGATGTTTGAAGACAGTGGCACTG 186  
 QY 1862 GCTATTTGGAAGCCGGAACGCTCCACCGGGAGTTAGCTGAGATGAGGGTTAGATGAATG 1921  
 Db 187 GCTATTTGGAAGCCGGAACGCTCCACCGGGAGTTAGCTGAGATGAGGGTTAGATGAATG 246  
 QY 1922 CCTGGGGTCAGAAAGCGCAGTATGTTTCTGAGAGTGTGCGCAACTGATGCTCTTG 1981  
 Db 247 CCTGGGGTCAGAAAGCGCAGTATGTTTCTGAGAGTGTGCGCAACTGATGCTCTTG 306  
 QY 1982 CAACGAGCAAGATCTGAGCAATTCATCAACTCAAGCTCTCAAGGTTCTTCCCCCAAGAA 2041  
 Db 307 CAACGAGCAAGATCTGAGCAATTCATCAACTCAAGGTTCTTCCCCCAAGAA 366  
 QY 2042 ATTGATATATGCTTTTATGTTTCTGATGGAATTTAAGTTTGTGCTCGGTAT 2101  
 Db 367 ATTGATATATGCTTTTATGTTTCTGATGGAATTTAAGTTTGTGCTCGGTAT 426  
 QY 2102 GCATCTGTTATGATATATATCTATGATTCATT 2132  
 Db 427 GCATCTGTTATGATATATATCTATGATTCATT 457

RESULT 6  
 AU227277/c 421 bp mRNA linear EST 23-APR-2002  
 LOCUS AU227277 RAFL14 Arabidopsis thaliana cDNA clone RAFL14-93-K05 3',  
 DEFINITION mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 421)  
 Seki,M., Narusaka,M., Ishida,J., Kamaya,A., Satou,M., Nakajima,M.,  
 Akiyama,K., Emju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J.,  
 Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinozaki,K.,  
 Muramatsu,M., Hayashizaki,Y. and Shinozaki,K.  
 Large scale analysis of Arabidopsis full-length cDNA  
 JOURNAL  
 TITLES  
 JOURNAL  
 Unpublished (2002)  
 Contact: Motoaki Seki  
 Plant Functional Genomics Research Group  
 RIKEN Genomic Sciences Center  
 3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-4359  
 Fax: 81-298-36-9060  
 Email: mseki@frc.riken.go.jp  
 An Arabidopsis full-length cDNA library was constructed essentially  
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI  
 and XhoI was ligated to modified lambda PLC-1 vector (Carninci et  
 al., submitted for publication) digested with BamHI and SalI. This  
 clone is in a modified plasmid vector. Please visit our web  
 site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further  
 details.

FEATURES  
 source  
 1..421  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3702"  
 /clone="RAFL14-93-K05"  
 /issue\_type="root"  
 /lab\_host="DH10B"  
 /clone\_1b="RAFL14"  
 /note="Site\_1: BamHI; Site\_2: SalI"



## ORIGIN

Query Match 10.0%; Score 329; DB 9; Length 380;  
 Best Local Similarity 99.7%; Pred. No. 3.1e-153;  
 Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

/strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="APD25d02R"  
 /issue\_type="aboveground organs"  
 /dev\_stage="two to six-week old"  
 /clone\_lib="Arabidopsis thaliana aboveground organs two to six-week old"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 8.9%; Score 290; DB 28; Length 290;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-133;  
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Class: BAC ends  
 High quality sequence stop: 290.  
 Location/Qualifiers  
 1..290  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="F27C8"  
 /sex="hermaphrodite"  
 /clone\_lib="IGF"  
 /note="Vector: BelosBACII; Site\_1: EcoRI; Site\_2: EcoRI;  
 Produced by Thomas Altmann"

## FEATURES

source

Class: BAC ends  
 High quality sequence stop: 290.  
 Location/Qualifiers  
 1..290  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="F27C8"  
 /sex="hermaphrodite"  
 /clone\_lib="IGF"  
 /note="Vector: BelosBACII; Site\_1: EcoRI; Site\_2: EcoRI;  
 Produced by Thomas Altmann"

QY 971 ATCCGAGAGGATTAACGCAATGTCGGGCACTTCAAGCTAACATCTGGTCCGGGACG 1030  
 DB 1 ATCCGAGAGGATTAACGCAATGTCGGGCACTTCAAGCTAACATCTGGTCCGGGACG 60  
 QY 1031 AGCGTTGAGCAAAATGATGATTAACAATTGGGCAACCCCACTGATCTCGCCCTCC 1090  
 DB 61 AGCGTTGAGCAAAATGATGATTAACAATTGGGCAACCCCACTGATCTCGCCCTCC 120  
 QY 1091 TTGGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1150  
 DB 121 TTGGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
 QY 1151 GGAATTTCTGACGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1210  
 DB 181 GGAATTTCTGACGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 QY 1211 TGCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1270  
 DB 241 TGCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 QY 1271 AAAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1330  
 DB 301 AAAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 QY 1331 TAGCTTTCATCGAGCAGA 1350  
 DB 361 TAGCTTTCATCGAGCAGA 380

RESULT 9  
 A0010650 290 bp DNA linear GSS 29-MAY-1998  
 LOCUS F27C8TIC IGF Arabidopsis thaliana genomic clone F27C8, genomic  
 DEFINITION survey sequence.  
 ACCESSION A0010650  
 VERSION A0010650.1 GI:3165927  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 290)  
 Roumley,S.D., Suh,E.J., Wible,C., Golden,K., Shateman,S., Choi,P.,  
 Yu,K., Akintetoye,B., Shen,K., Goomasekaram,S., Miltscher,J.,  
 Adams,M.D. and Venter,J.C.  
 A BAC End Sequence Database for Identifying Minimal Overlaps in  
 Arabidopsis Genomic Sequencing. Update 4  
 Unpublished (1998)  
 Other GSSs: F27C8TFC  
 Contact: Steve Roumley  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 0208  
 Email: roumley@tigr.org  
 Seq primer: M13 Reverse

QY 1668 TTCTGCTTTTACTTCTTTAATTTTCTCTTGATTTCTACTGATTTTGAATGTTACTT 1727  
 DB 1 TTCTGCTTTTACTTCTTTAATTTTCTCTTGATTTCTACTGATTTTGAATGTTACTT 60  
 QY 1728 GTAGTGGCTGGCATGGGCAACCAAGAGCTGCTGAATCTTCAAGATAGAGGCTCT 1787  
 DB 61 GTAGTGGCTGGCATGGGCAACCAAGAGCTGCTGAATCTTCAAGATAGAGGCTCT 120  
 QY 1788 TAGAGCAGCCCATTTCTTAATGATCACAGGCGCATGCTGGATGATGATGATGATGATGAT 1847  
 DB 121 TAGAGCAGCCCATTTCTTAATGATCACAGGCGCATGCTGGATGATGATGATGATGATGAT 180  
 QY 1848 GAGCAGTCCCATGCTGATTTTGAAGCCCAACCCCTCCACCGGAGTATGATGATGATGATGAT 1907  
 DB 181 GAGCAGTCCCATGCTGATTTTGAAGCCCAACCCCTCCACCGGAGTATGATGATGATGATGAT 240  
 QY 1908 GTTAGATGAAATTGGCTGGGGTGGAGAGCGCAGTATGTTTTCGAGGTG 1957  
 DB 241 GTTAGATGAAATTGGCTGGGGTGGAGAGCGCAGTATGTTTTCGAGGTG 290

## RESULT 10

A0959659 523 bp DNA linear GSS 28-JAN-2000  
 LOCUS LEREM21LR LERB Arabidopsis thaliana genomic clone LEREM21, genomic  
 DEFINITION survey sequence.  
 ACCESSION A0959659  
 VERSION A0959659.1 GI:6787360  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 523)  
 Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Uteerbach,T.,  
 Feldblyum,T., Liang,F., Creasy,T. and Frazer,C.M.  
 Genomic survey sequencing of Landsberg erecta ecotype of  
 Arabidopsis thaliana and identification of sequence-based  
 polymorphisms  
 Unpublished (2000)  
 Contact: Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 0208  
 Email: al@tigr.org  
 For additional information, see <http://www.tigr.org/cdb/at/at.html>  
 Seq primer: TR  
 Class: shotgun.  
 Location/Qualifiers  
 1..523  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"

## FEATURES

source

Class: BAC ends  
 High quality sequence stop: 290.  
 Location/Qualifiers  
 1..523  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"



/strain="LANDSBERG\_ERECTA"  
 /db\_xref="taxon:3702"  
 /clone="LBERM21"  
 /clone\_lib="LBERF"  
 /note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was  
 sheared to 0.6-0.8 Kbp before ligation."

## ORIGIN

Query Match 8.3%; Score 273; DB 28; Length 523;  
 Best Local Similarity 99.3%; Pred. No. 3.8e-125;

Matches 423; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 862 GAAACACTTCTGGAAAACTTGGGTTCTCAGAAATTCATCTCTAGAGCTTGGGGTG 921  
 DB 25 GAAACACTTCTGGAAAACTTGGGTTCTCAGAAATTCATCTCTAGAGCTTGGGGTG 84  
 QY 922 GTCAGCAGCAAGGAGGAGGTAGCAACGTATCTGGAGAGAAACAATGTATCCGGAGAG 981  
 DB 85 GTCAGCAGCAAGGAGGAGGTAGCAACGTATCTGGAGAGAAACAATGTATCCGGAGAG 144  
 QY 982 GTAAAGCGCAATGCTGGGGCATTCAGCTAACATCTCTGGGGAGCAAGCGTTAGCA 1041  
 DB 145 GTAAAGCGCAATGCTGGGGCATTCAGCTAACATCTCTGGGGAGCAAGCGTTAGCA 204  
 QY 1042 GAAAGTATGATACAACTTGTGGCAACCCCACTGTATCTCGCCCTCTTGGAGAGAG 1101  
 DB 205 GAAAGTATGATACAACTTGTGGCAACCCCACTGTATCTCGCCCTCTTGGAGAGAG 264  
 QY 1102 GATGAATTTGGCAGGCAAGAGAGGTTCTGCTCAGACACAGCTGTGAGAGATTTCTTG 1161  
 DB 265 GATGAATTTGGCAGGCAAGAGAGGTTCTGCTCAGACACAGATGTGAGAGATTTCTTG 324  
 QY 1162 AGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1221  
 DB 325 AGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 384  
 QY 1222 ATTCTGATGACGACCTTGGCAAGTATGATGATGATGATGATGATGATGATGATGATG 1281  
 DB 385 ATTCTGATGACGACCTTGGCAAGTATGATGATGATGATGATGATGATGATGATGATG 444  
 QY 1282 GATCAC 1287  
 DB 445 GATCAC 450

## RESULT 11

AL945644 536 bp DNA linear GSS 30-NOV-2002  
 LOCUS AL945644/C Arabidopsis thaliana T-DNA flanking sequence GK-290H04-015353,  
 DEFINITION genomic survey sequence.

ACCESSION AL945644  
 VERSION AL945644.1 GI:24402266  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Bacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

## REFERENCE

AUTHORS Strizhov, N., Li, Y., Rosso, M., Vlehoever, P., Dekker, K., Saedler, H.  
 and Weishaar, B.  
 TITLE A pipeline for automated high-throughput generation of FSTs  
 (flanking sequence tags) from Arabidopsis thaliana T-DNA  
 transformed lines  
 JOURNAL Unpublished

## AUTHORS

REFERENCE Rosso, M., Strizhov, N., Li, Y., Reies, B., Dekker, K. and Weishaar, B.  
 A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)  
 for flanking sequence tag based reverse genetics

JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 536)  
 AUTHORS Li, Y., Rosso, M., Strizhov, N. and Weishaar, B.  
 TITLE Direct Submission

## JOURNAL

Submitted (21-OCT-2002) Weishaar B., Max-Planck-Institut fuer  
 Zuechtungsforshung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany  
 This sequence is recovered from the left border of the T-DNA. It  
 indicates an insertion within the locus defined by clone MQM1. The  
 sequences are generated at the MPI for Plant Breeding Research in  
 the context of the GABI-Kat project. GABI-Kat is part of the German  
 Plant Genomics program designated 'GABI'. Information on line  
 availability can be found at:  
 http://www.mpiz-koeln.mpg.de/GABI-Kat/.

## FEATURES

source

Location/Qualifiers  
 1..536

/organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone="GK-290H04-015353"  
 /clone\_lib="Arabidopsis thaliana T-DNA insertion lines"  
 /note="PCR was performed on DNA from Arabidopsis thaliana  
 plants (T1) which were transformed with the T-DNA from  
 vector pAC161. The lines contain one or more T-DNA  
 insertions. The DNA fragment(s) resulting from the PCR  
 were directly sequenced to determine the genomic sequence  
 flanking the insertion. Sequences displaying significant  
 similarity to the A. thaliana nuclear genome sequence were  
 processed for submission. T-DNA derived sequences were  
 removed."

## ORIGIN

Query Match 7.1%; Score 232; DB 29; Length 536;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-104;  
 Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3044 TGGGCAATCTAAGAGATTAATGATTAATTCCTCCAAATTTAGTAGAGGATCTAA 3103  
 DB 527 TGGGCAATCTAAGAGATTAATGATTAATTCCTCCAAATTTAGTAGAGGATCTAA 468  
 QY 3104 GGAAGCATTAAGTTCTTGTGACTAAACCAAGTTCTTGTGATTTTGTGTTTGGT 3163  
 DB 467 GGAAGCATTAAGTTCTTGTGACTAAACCAAGTTCTTGTGATTTTGTGTTTGGT 408  
 QY 3164 AAAATTCTATGAAAGTTAGACATATTACCAAGCTCAGAGTGAATCAGAAATGGCA 3223  
 DB 407 AAAATTCTATGAAAGTTAGACATATTACCAAGCTCAGAGTGAATCAGAAATGGCA 348  
 QY 3224 ATCAAAATGATGTTTGTGATTAATTTATCTCAAAATTAATGGTAAAT 3275  
 DB 347 ATCAAAATGATGTTTGTGATTAATTTATCTCAAAATTAATGGTAAAT 296

## RESULT 12

AV566465 619 bp mRNA linear EST 07-SEP-2000  
 LOCUS AV566465 Arabidopsis thaliana green siliques Columbia Arabidopsis  
 DEFINITION thaliana cDNA clone SQ244b06f 3', mRNA sequence..

ACCESSION AV566465  
 VERSION AV566465.1 GI:8737917  
 KEYWORDS EST.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Bacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE Asamizu, E., Nakamura, Y., Seto, S. and Tabata, S.  
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries

## JOURNAL

REFERENCE 20363093  
 MEDLINE 10907847  
 PUBMED 10907847

COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute

Yana 1532-3, Kisearazu, Chiba 292-0812, Japan  
Email: asamiizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.  
Location/Qualifiers

1. 619  
/organism="Arabidopsis thaliana"  
/mol\_type="RNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="SQ244b6F"  
/issue\_type="green siliques"  
/clone\_1b="Arabidopsis thaliana green siliques Columbia"  
/note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2: XhoI"

## ORIGIN

Query Match 6.4%; Score 210; DB 9; Length 619;  
Best Local Similarity 100.0%; Pred. No. 1.6e-93;  
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2170 CAAGAGATGGTTGAAGAGCTGAGCGAGATCTCTGAGACATACAGAGCTGAATAC 2229  
DB 619 CAGAGATGTTTGAAGAGCTGAGCGAGATCTCTGAGACATACAGAGCTGAATAC 560  
QY 2230 TTTAAGACCAAGCTCTCAAAACGACAGACAGCCAGAGCTTGAAGATCTCTGAA 2289  
DB 559 TTTAAGACCAAGCTCTCAAAACGACAGACAGCCAGAGCTTGAAGATCTCTGAA 500  
QY 2290 ATTATGACGAGAGCTGCTGAGAGCTGAGAGATATGATCTGAGAGAGAGACT 2349  
DB 499 ATTATGACGAGAGCTGCTGAGAGCTGAGAGATATGATCTGAGAGAGAGACT 440  
QY 2350 AAGATCGAGCTGAACAGACAGAGAGAG 2379  
DB 439 AAGATCGAGCTGAACAGACAGAGAGAG 410

RESULT 13 443 bp DNA linear GSS 30-JAN-2002  
BH618393/c Arabidopsis thaliana T-DNA insertion line Arabidopsis  
LOCUS SALK\_039005 Arabidopsis thaliana T-DNA insertion line Arabidopsis  
DEFINITION thaliana genomic clone SALK\_039005, genomic survey sequence.

ACCESSION BH618393.1 GI:18428488  
VERSION BH618393  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 443)  
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,  
Gadgil, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,  
Shim, P., Zimmerman, J. and Ecker, J.R.  
A sequence-indexed library of insertion mutations in the  
Arabidopsis genome  
Unpublished (2001)

JOURNAL CONTACT: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: eckere@salil.salk.edu

COMMENT This is single pass sequence recovered from the left border of  
T-DNA. This sequence lies within an annotated intron of At5g23570.  
Class: T-DNA tagged.  
Location/Qualifiers

1. 443  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="SALK\_039005"

## ORIGIN

Query Match 5.6%; Score 184; DB 28; Length 443;  
Best Local Similarity 100.0%; Pred. No. 1.9e-80;  
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1694 CTCTTGATCTTACGATCTTGAAGATGTTTACATTTAGTGGCTGGCATGGCAACCAAG 1753  
DB 443 CTCTTGATCTTACGATCTTGAAGATGTTTACATTTAGTGGCTGGCATGGCAACCAAG 384  
QY 1754 AGCTGCGAATATCTTGAAGATGTTTACATTTAGTGGCTGGCATGGCAACCAAG 1813  
DB 383 AGCTGCGAATATCTTGAAGATGTTTACATTTAGTGGCTGGCATGGCAACCAAG 324  
QY 1814 AGGCGCATCTGAGATGTTTACATTTAGTGGCTGGCATGGCAACCAAG 1873  
DB 323 AGGCGCATCTGAGATGTTTACATTTAGTGGCTGGCATGGCAACCAAG 264  
QY 1874 CCGA 1877  
DB 263 CCGA 260

RESULT 14 395 bp DNA linear GSS 02-JUL-2003  
BX547725 Arabidopsis thaliana T-DNA flanking sequence GK-547E01-020631,  
LOCUS genomic survey sequence.  
DEFINITION BX547725  
ACCESSION BX547725.1 GI:32440534  
VERSION BX547725  
KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 395)  
Strizhov, N., Li, Y., Rosso, M., Vlehever, P., Dekker, K., Saedler, H.  
and Weishaar, B.  
A pipeline for automated high-throughput generation of FSTs  
(flanking sequence tags) from Arabidopsis thaliana T-DNA  
transformed lines  
Unpublished

JOURNAL CONTACT: Strizhov, N., Li, Y., Rosso, M., Dekker, K. and Weishaar, B.  
Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weishaar, B.  
A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)  
for flanking sequence tag based reverse genetics  
Unpublished  
3 (bases 1 to 395)

JOURNAL CONTACT: Strizhov, N., Li, Y., Rosso, M. and Weishaar, B.  
Submitted (01-JUL-2003) Weishaar, B., Max-Planck-Institut fuer  
Zoechnungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany  
This sequence is recovered from the left border of the T-DNA. It  
indicates an insertion within the locus defined by clone M0M1. The  
sequences are generated at the GABI-Kat project. GABI-Kat is part of the German  
Plant Genomics program designated 'GABI'. Information on line  
availability can be found at:  
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

COMMENT This is single pass sequence recovered from the left border of  
T-DNA. This sequence lies within an annotated intron of At5g23570.  
Class: T-DNA tagged.  
Location/Qualifiers

1. 395  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"

/clone="GK-547E01-020631"  
 /clone\_lib="Arabidopsis thaliana T-DNA insertion lines"  
 /note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

## ORIGIN

Query Match 5.2%; Score 170; DB 29; Length 395;  
 Best Local Similarity 99.5%; Pred. No. 2.1e-73;  
 Matches 220; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 163 AGCTTAATAAGCTTCCTCATTTGCTCTTCTTGTGTCAGTTATTTCTTCTCCGAGATC 222  
 |||  
 DB 50 AGCTTAATAAGCTTCCTCATTTGCTCTTCTTGTGTCAGTTATTTCTTCTCCGAGATC 109  
 |||  
 QY 223 CTGACCTACTACTCTGCTCTGCTGCGGCTTAACTTACCTTCTCGTCTTACTCTGT 282  
 |||  
 DB 110 CTGACCTACTACTCTGCTCTGCTGCGGCTTAACTTACCTTCTCGTCTTACTCTGT 169  
 |||  
 QY 283 AAGTTTTCGCTTGAAGCTCCGATCGCTCACCGCATTCATTTGTCGATTTCTC 342  
 |||  
 DB 170 AAGTTTTCGCTTGAAGCTCCGATCGCTCACCGCATTCATTTGTCGATTTCTC 229  
 |||  
 QY 343 TTTTCTTCTGCTGAAAAATTTGCCCTAATGTTCTCGATTTTC 383  
 |||  
 DB 230 TTTTCTTCTGCTGAAAAATTTGCCCTAATGTTCTCGATTTTC 270  
 |||

## RESULT 15

BH169457

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BH169457 197 bp DNA linear GSS 03-OCT-2001  
 SALK\_001394 Arabidopsis thaliana TDNA insertion lines Arabidopsis  
 thaliana genomic clone SALK\_001394, genomic survey sequence.  
 BH169457  
 BH169457.1 GI:15904832  
 GSS.  
 Arabidopsis thaliana (thale cress)  
 Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 197)  
 Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,  
 Gadrihab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,  
 Shin,P., Zimmerman,J. and Ecker,J.R.  
 A Sequence-Indexed Library of Insertion Mutations in the  
 Arabidopsis Genome  
 Unpublished (2001)  
 Contact: Joseph R. Ecker  
 Salk Institute Genomic Analysis Laboratory (SIGAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6379  
 Email: ecker@salk.edu  
 This is single pass sequence recovered from the left border of  
 TDNA.  
 Class: TDNA tagged.  
 Location/Qualifiers  
 1..197  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone="SALK\_001394"  
 /clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
 /note="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion  
 elements. The resultant fragment for each line was  
 directly sequenced to determine the genomic sequence at  
 the site of insertion. Details of the protocols used can  
 be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)

## ORIGIN

Query Match 4.7%; Score 153; DB 28; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-65;  
 Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 GTTATAGGCTTGAAGCTTGAACAGTTGTTCAAGTTTGCGAGGACGAGACTGCTTCTT 804  
 |||  
 DB 1 GTTATAGGCTTGAAGCTTGAACAGTTGTTCAAGTTTGCGAGGACGAGACTGCTTCTT 60  
 |||  
 QY 805 CACAAGATGATGAGAGAGAGTGGAGTCATTTCCAGAAGAACAAAGAACAGAGAA 864  
 |||  
 DB 61 CACAAGATGATGAGAGAGAGTGGAGTCATTTCCAGAAGAACAAAGAACAGAGAA 120  
 |||  
 QY 865 ACACTTCTGAAAAAACTTGCGTTTCTCAGAAAT 897  
 |||  
 DB 121 ACACTTCTGAAAAAACTTGCGTTTCTCAGAAAT 153  
 |||

Search completed: April 9, 2004, 08:16:05  
 Job time : 5513.97 secs

**This Page Blank (uspto)**



TYPE: DNA  
ORGANISM: Escherichia coli  
US-09-736-734-1

Query Match 0.6%; Score 20; DB 4; Length 654;  
Best Local Similarity 100.0%; Pred. No. 7.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3164 AAAATTCATGAAAGTTA 3183  
|||||  
Db 153 AAAATTCATGAAAGTTA 134

RESULT 3  
US-09-615-192A-109/c

Sequence 109, Application US/09615192A  
Patent No. 6410718  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N.  
APPLICANT: Havukkala, Ilkka  
TITLE OF INVENTION: Materials and Methods for the  
FILE REFERENCE: 11000.1003C4U  
CURRENT APPLICATION NUMBER: US/09/615,192A  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 08/975,316  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: US 08/713,000  
PRIOR FILING DATE: 1996-09-11  
PRIOR APPLICATION NUMBER: US 09/169,789  
PRIOR FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 405  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 109  
LENGTH: 481  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-615-192A-109

Query Match 0.6%; Score 19; DB 4; Length 481;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTCAAGTTGGCAGG 787  
|||||  
Db 470 TGGTCAAGTTGGCAGG 452

RESULT 4  
US-09-169-789-109/c

Sequence 109, Application US/09169789  
Patent No. 6653528  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N.  
APPLICANT: Havukkala, Ilkka  
TITLE OF INVENTION: Materials and Methods for the  
FILE REFERENCE: 11000.1003C2  
CURRENT APPLICATION NUMBER: US/09/169,789  
PRIOR FILING DATE: 1998-10-09  
PRIOR APPLICATION NUMBER: US 08/975,316  
PRIOR FILING DATE: 1997-11-21  
PRIOR APPLICATION NUMBER: US 08/713,000  
PRIOR FILING DATE: 1996-09-11  
NUMBER OF SEQ ID NOS: 185  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 109  
LENGTH: 481  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-169-789-109

Query Match 0.6%; Score 19; DB 4; Length 481;

Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTCAAGTTGGCAGG 787  
|||||  
Db 470 TGGTCAAGTTGGCAGG 452

RESULT 5  
US-08-975-316-27/c

Sequence 27, Application US/08975316  
Patent No. 5952486  
GENERAL INFORMATION:  
APPLICANT: BLOKSBERG, Leonard N.; HAVUKKALA, Ilkka  
APPLICANT: and GRIERSON, Alastair W.  
TITLE OF INVENTION: MATERIALS AND METHODS FOR  
THE MODIFICATION OF PLANT LIGNIN CONTENT  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Law Offices of Ann W. Speckman  
STREET: 2601 Elliott Avenue, Suite 4185  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98121  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/975,316  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/713,000  
FILING DATE: September 11, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: SLEATH, Janet  
REGISTRATION NUMBER: 37,007  
REFERENCE/DOCKET NUMBER: 11000/1003C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-269-0565  
TELEFAX: 206-269-0563  
TELEX:  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 495 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-975-316-27

Query Match 0.6%; Score 19; DB 2; Length 495;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTCAAGTTGGCAGG 787  
|||||  
Db 484 TGGTCAAGTTGGCAGG 466

RESULT 6  
US-09-615-192A-27/c

Sequence 27, Application US/09615192A  
Patent No. 6410718  
GENERAL INFORMATION:  
APPLICANT: Bloksberg, Leonard N.  
APPLICANT: Havukkala, Ilkka  
TITLE OF INVENTION: Materials and Methods for the  
FILE REFERENCE: 11000.1003C4U  
CURRENT APPLICATION NUMBER: US/09/615,192A

;; CURRENT FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 08/975,316  
;; PRIOR FILING DATE: 1997-11-21  
;; PRIOR APPLICATION NUMBER: US 08/713,000  
;; PRIOR FILING DATE: 1996-09-11  
;; PRIOR APPLICATION NUMBER: US 09/169,789  
;; PRIOR FILING DATE: 1998-10-09  
;; NUMBER OF SEQ ID NOS: 405  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO: 27  
;; LENGTH: 495  
;; TYPE: DNA  
;; ORGANISM: Eucalyptus grandis  
US-09-615-192A-27

Query Match 0.6%; Score 19; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTTCAGGTTGGCAGG 787  
DB 484 TGGTTCAGGTTGGCAGG 466

RESULT 7  
US-09-169-789-27/c  
; Sequence 27, Application US/09169789  
; Patent No. 6653528  
; GENERAL INFORMATION:  
; APPLICANT: Bloksberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; FILE REFERENCE: 11000.1003c2  
; CURRENT APPLICATION NUMBER: US/09/169,789  
; CURRENT FILING DATE: 1998-10-09  
; EARLIER APPLICATION NUMBER: US 08/975,316  
; EARLIER FILING DATE: 1997-11-21  
; EARLIER APPLICATION NUMBER: US 08/713,000  
; EARLIER FILING DATE: 1996-09-11  
; NUMBER OF SEQ ID NOS: 185  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO: 27  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-09-169-789-27

Query Match 0.6%; Score 19; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTTCAGGTTGGCAGG 787  
DB 484 TGGTTCAGGTTGGCAGG 466

RESULT 8  
US-08-484-105-19  
; Sequence 19, Application US/08484105  
; Patent No. 5589341  
; GENERAL INFORMATION:  
; APPLICANT: STILLMAN, Bruce  
; APPLICANT: BELL, Stephen P  
; APPLICANT: KOBAYASHI, Ryuji  
; APPLICANT: RINE, Jasper  
; APPLICANT: MCNALLY, Margit  
; APPLICANT: MCNALLY, Francis J  
; APPLICANT: LAURENSEN, Patricia  
; APPLICANT: HERSKOWITZ, Ira  
; APPLICANT: LI, Joachim J  
; APPLICANT: GAVIN, Kimberly  
; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

;; NUMBER OF SEQUENCES: 24  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
;; STREET: 4 Embarcadero Center, Suite 3400  
;; CITY: San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94111-4187  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/484,105  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Osman Ph.D., Richard Aron  
;; REGISTRATION NUMBER: 36,627  
;; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 494-8700  
;; TELEFAX: (415) 494-8771  
;; TELEX: 910 277299  
;; INFORMATION FOR SEQ ID NO: 19:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1480 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 277..1365  
US-08-484-105-19

Query Match 0.6%; Score 19; DB 1; Length 1480;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1655 TAAGGTGAATTCCTCTGT 1673  
DB 238 TAAGGTGAATTCCTCTGT 256

RESULT 9  
US-08-484-106-19  
; Sequence 19, Application US/08484106  
; Patent No. 5614618  
; GENERAL INFORMATION:  
; APPLICANT: STILLMAN, Bruce  
; APPLICANT: BELL, Stephen P  
; APPLICANT: KOBAYASHI, Ryuji  
; APPLICANT: RINE, Jasper  
; APPLICANT: MCNALLY, Margit  
; APPLICANT: MCNALLY, Francis J  
; APPLICANT: LAURENSEN, Patricia  
; APPLICANT: HERSKOWITZ, Ira  
; APPLICANT: LI, Joachim J  
; APPLICANT: GAVIN, Kimberly  
; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible



```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,106
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/FAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1480 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 277..1365
US-08-484-106-19

```

```

Query Match      0.6%; Score 19; DB 1; Length 1480;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1655 TAAGTGAATTCCTCTGT 1673
DB      238 TAAGTGAATTCCTCTGT 256

```

```

RESULT 10
US-09-377-557-9/c
Sequence 9, Application US/09377557
Patent No. 6297055
GENERAL INFORMATION:
APPLICANT: Falco, S. Carl
APPLICANT: Famodu, Layo O.
APPLICANT: Orozco, Emil M. Jr.
TITLE OF INVENTION: Amino Acid Decarboxylases
FILE REFERENCE: BB-1237
CURRENT APPLICATION NUMBER: US/09/377,557
CURRENT FILING DATE: 1999-08-19
EARLIER APPLICATION NUMBER: 60/099,493
EARLIER FILING DATE: September 8, 1998
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Microsoft Office 97
SEQ ID NO 9
LENGTH: 1857
TYPE: DNA
ORGANISM: Oryza sativa
US-09-377-557-9

```

```

Query Match      0.6%; Score 19; DB 3; Length 1857;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3152 TTTTGTGTAATTT 3170
DB      1844 TTTTGTGTAATTT 1826

```

```

RESULT 11
US-09-041-075A-4
Sequence 4, Application US/09041075A
Patent No. H002022
GENERAL INFORMATION:
APPLICANT: Heidler, Steven A
APPLICANT: Radding, Jeffrey A

```

```

TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
FILE REFERENCE: X-11242 Sequence List
Patent No. H002022
CURRENT APPLICATION NUMBER: US/09/041,075A
CURRENT FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/043,591
PRIOR FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 2000
TYPE: DNA
ORGANISM: Candida krusei
FEATURE:
NAME/KEY: CDS
LOCATION: (300)..(1739)
US-09-041-075A-4

```

```

Query Match      0.6%; Score 19; DB 1; Length 2000;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1218 GATGATTCGATGACGACC 1236
DB      1560 GATGATTCGATGACGACC 1578

```

```

RESULT 12
US-09-041-075A-6
Sequence 6, Application US/09041075A
Patent No. H002022
GENERAL INFORMATION:
APPLICANT: Heidler, Steven A
APPLICANT: Radding, Jeffrey A
TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
FILE REFERENCE: X-11242 Sequence List
Patent No. H002022
CURRENT APPLICATION NUMBER: US/09/041,075A
CURRENT FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/043,591
PRIOR FILING DATE: 1997-04-15
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 2000
TYPE: RNA
ORGANISM: Candida krusei
US-09-041-075A-6

```

```

Query Match      0.6%; Score 19; DB 1; Length 2000;
Best Local Similarity 73.7%; Pred. No. 25;
Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1218 GATGATTCGATGACGACC 1236
DB      1560 GATGATTCGATGACGACC 1578

```

```

RESULT 13
US-09-595-684B-30
Sequence 30, Application US/09595684B
Patent No. 6544766
GENERAL INFORMATION:
APPLICANT: Beraud, Christophe
APPLICANT: Ohashi, Cara
APPLICANT: Sakowicz, Roman
APPLICANT: Valsberg, Eugeni
APPLICANT: Wood, Kenneth
APPLICANT: Yu, Ming
TITLE OF INVENTION: Human kinesins and methods of producing
TITLE OF INVENTION: and purifying human kinesins
FILE REFERENCE: cytop036
CURRENT APPLICATION NUMBER: US/09/595,684B

```

;; CURRENT FILING DATE: 2002-06-24  
;; PRIOR APPLICATION NUMBER: 09/295,612  
;; PRIOR FILING DATE: 2000-04-20  
;; NUMBER OF SEQ ID NOS: 105  
;; SOFTWARE: FastSeq for windows Version 4.0  
;; SEQ ID NO 30  
;; LENGTH: 8257  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-595-684B-30

Query Match 0.6%; Score 19; DB 4; Length 8257;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2873 TTGATCTGGAGAAAGATT 2891  
DB 2141 TTGATCTGGAGAAAGATT 2159

RESULT 14  
US-09-620-312D-130  
;; Sequence 130, Application US/09620312D  
;; Patent No. 6569662  
;; GENERAL INFORMATION:  
;; APPLICANT: Tang, Y. Tom  
;; APPLICANT: Liu, Chenghua  
;; APPLICANT: Asundi, Vinod  
;; APPLICANT: Zhang, Jie  
;; APPLICANT: Ren, Feiyun  
;; APPLICANT: Chen, Rui-hong  
;; APPLICANT: Zhao, Qing A.  
;; APPLICANT: Mehrman, Tom  
;; APPLICANT: Xue, Aidong J.  
;; APPLICANT: Yang, Yonghong  
;; APPLICANT: Wang, Jian-Rui  
;; APPLICANT: Zhou, Ping  
;; APPLICANT: Ma, Yundong  
;; APPLICANT: Wang, Dunrui  
;; APPLICANT: Wang, Zhiwei  
;; APPLICANT: John Tillinghast  
;; APPLICANT: Drmanac, Radoje T.  
;; TITLE OF INVENTION: No. 6569662zel Nucleic Acids and  
;; FILE REFERENCE: 784CIP28  
;; CURRENT APPLICATION NUMBER: US/09/620,312D  
;; CURRENT FILING DATE: 2000-07-19  
;; PRIOR APPLICATION NUMBER: 09/552,317  
;; PRIOR FILING DATE: 2000-04-25  
;; PRIOR APPLICATION NUMBER: 09/488,725  
;; PRIOR FILING DATE: 2000-01-21  
;; NUMBER OF SEQ ID NOS: 1105  
;; SOFTWARE: pc\_fl\_genes Version 1.0  
;; SEQ ID NO 130  
;; LENGTH: 8503  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (91)..(8082)  
US-09-620-312D-130

Query Match 0.6%; Score 19; DB 4; Length 8503;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2873 TTGATCTGGAGAAAGATT 2891  
DB 2141 TTGATCTGGAGAAAGATT 2159

RESULT 15  
US-09-163-748C-3/c

;; Sequence 3, Application US/09163748C  
;; Patent No. 6509172  
;; GENERAL INFORMATION:  
;; APPLICANT: DeBacker, Oliver  
;; APPLICANT: Van den Eynde, Benoit  
;; APPLICANT: Boon-Fallieur, Thierry  
;; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode Member Of The Gage  
;; FILE REFERENCE: LUD 5558  
;; CURRENT APPLICATION NUMBER: US/09/163,748C  
;; CURRENT FILING DATE: 1998-09-30  
;; NUMBER OF SEQ ID NOS: 27  
;; SEQ ID NO 3  
;; LENGTH: 9531  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: 92,232,1041,7412,9038-9040  
;; OTHER INFORMATION: Identity of several nucleotides not known  
US-09-163-748C-3

Query Match 0.6%; Score 19; DB 4; Length 9531;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 TACAAATGAAACAAA 109  
DB 7045 TACAAATGAAACAAA 7027

Search completed: April 9, 2004, 08:20:42  
Job time : 165.43 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using bw model

Run on: April 9, 2004, 05:52:11 ; Search time 923.457 Seconds  
(without alignments)  
13304.392 Million cell updates/sec

Title: US-10-030-829-1  
Perfect score: 3275  
Sequence: 1 gcaacaacaacaataa.....caaatatagggatacaat 3275

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2475585 seqs, 1875730760 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4951170

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	0.7	6136	US-10-257-166-123	Sequence 123, App
2	22	0.7	9888	US-10-311-455-1214	Sequence 1214, Ap
3	21	0.6	2164	US-10-027-613-250839	Sequence 250839,
4	21	0.6	6048	US-10-311-455-482	Sequence 482, App
5	21	0.6	6072	US-10-311-455-4	Sequence 4, Appli
6	21	0.6	8245	US-10-221-714A-170	Sequence 170, App
7	20	0.6	476	US-10-001-843-87	Sequence 87, Appl
8	20	0.6	654	US-09-736-734-1	Sequence 1, Appli
9	20	0.6	654	US-10-396-911-1	Sequence 133, App
10	20	0.6	1023	US-10-074-475-133	Sequence 91177, A
11	20	0.6	1377	US-10-424-599-91177	Sequence 10949, A
12	20	0.6	1461	US-10-282-122A-10949	Sequence 91176, A
13	20	0.6	1589	US-10-424-599-91176	Sequence 1756, Ap
14	20	0.6	2418	US-10-108-260A-1756	Sequence 1, Appli
15	20	0.6	2985	US-09-947-953-1	Sequence 1, Appli

16	20	0.6	5278	12	US-10-221-714A-98	Sequence 98, Appl
17	20	0.6	5278	14	US-10-311-455-796	Sequence 796, App
18	20	0.6	5376	14	US-10-311-455-2123	Sequence 2123, Ap
19	20	0.6	5453	12	US-10-221-613-176	Sequence 376, App
20	20	0.6	6077	14	US-10-311-455-1828	Sequence 1828, Ap
21	20	0.6	6212	14	US-10-240-453-7	Sequence 7, Appli
22	20	0.6	6412	12	US-10-221-714A-353	Sequence 353, App
23	20	0.6	6412	14	US-10-311-455-1883	Sequence 1883, Ap
24	20	0.6	7624	14	US-10-311-455-2085	Sequence 2085, Ap
25	20	0.6	10696	14	US-10-311-455-291	Sequence 291, App
26	20	0.6	10696	14	US-10-240-452-39	Sequence 39, Appl
27	20	0.6	11422	14	US-10-311-455-191	Sequence 191, App
28	20	0.6	11422	16	US-10-257-166-17	Sequence 1793, Ap
29	20	0.6	13732	14	US-10-311-455-1793	Sequence 1793, Ap
30	20	0.6	14708	12	US-10-221-714A-499	Sequence 499, App
31	20	0.6	14708	14	US-10-239-676-221	Sequence 221, App
32	20	0.6	14708	14	US-10-311-455-2217	Sequence 2217, Ap
33	20	0.6	14708	14	US-10-240-453-323	Sequence 323, App
34	20	0.6	16724	14	US-10-311-455-1064	Sequence 1064, Ap
35	20	0.6	16724	14	US-10-240-485-90	Sequence 90, Appl
36	20	0.6	17580	14	US-10-240-452-12	Sequence 12, Appl
37	20	0.6	17993	9	US-09-768-781-5	Sequence 5, Appli
38	20	0.6	34769	12	US-10-221-714A-502	Sequence 502, App
39	20	0.6	73334	14	US-10-311-455-2098	Sequence 2098, Ap
40	20	0.6	302603	12	US-10-271-416-8	Sequence 8, Appli
41	20	0.6	3673778	14	US-10-312-841-1	Sequence 1, Appli
42	20	0.6	3673778	14	US-10-312-841-2	Sequence 2, Appli
43	19	0.6	172	12	US-10-085-783A-19246	Sequence 19246, A
44	19	0.6	172	15	US-10-242-535A-19246	Sequence 19246, A
45	19	0.6	385	14	US-10-066-543-1881	Sequence 1881, Ap

## ALIGNMENTS

RESULT 1  
US-10-257-166-123  
; Sequence 123, Application US/10257166  
; Publication No. US20040023230A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBERCK, Christian  
; APPLICANT: BERLIN, Kutt  
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of  
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics  
; FILE REFERENCE: 5013.1011  
; CURRENT APPLICATION NUMBER: US/10/257,166  
; CURRENT FILING DATE: 2002-10-07  
; PRIOR APPLICATION NUMBER: PCT/EP01/07470  
; DE 10032529.7  
; DE 10043826.1  
; PRIOR FILING DATE: 2001-06-29  
; 2000-06-30  
; 2000-09-01  
; NUMBER OF SEQ ID NOS: 178  
; SEQ ID NO 123  
; LENGTH: 6136  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-257-166-123

Query Match 0.7%; Score 22; DB 16; Length 6136;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3145 TATTTGTTTTTTTGTGTA 3166  
|||||  
Db 2122 TATTTGTTTTTTTGTGTA 2143  
|||||

RESULT 2

```
US-10-311-455-1214
; Sequence 1214, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1214
; LENGTH: 9888
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1214

Query Match
Best Local Similarity 0.7%; Score 22; DB 14; Length 9888;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3142 TAGTATTTTGTGTTTTTGTGTT 3163
DB 8219 TAGTATTTTGTGTTTTTGTGTT 8240

RESULT 3
US-10-027-632-250839
; Sequence 250839, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 250839
; LENGTH: 2164
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-250839

Query Match
Best Local Similarity 0.6%; Score 21; DB 15; Length 2164;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 3154 TTTTGTGTAATTTTCATA 3174
DB 842 TTTTGTGTAATTTTCATA 862

RESULT 4
US-10-311-455-482
; Sequence 482, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 482
; LENGTH: 6048
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-482

Query Match
Best Local Similarity 0.6%; Score 21; DB 14; Length 6048;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 526 TTGCTTTTGTGTTTTGTTT 546
DB 5395 TTGCTTTTGTGTTTTGTTT 5415

RESULT 5
US-10-311-455-4
; Sequence 4, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 4
; LENGTH: 6072
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-4

Query Match
Best Local Similarity 0.6%; Score 21; DB 14; Length 6072;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	3145	TATTTTGTTTTTTTTGCTAA	3165
Dd	3717	TATTTTGTTTTTTTTGCTAA	3737

```

RESULT 6
US-10-221-714A-170
Sequence: 170, Application US/10221714A
Publication No. US20040046254A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with
TITLE OF INVENTION: tumor suppressor genes and oncogenes
FILE REFERENCE: 5013.1005
CURRENT APPLICATION NUMBER: US/10/221,714A
CURRENT FILING DATE: 2003-01-21
PRIORITY APPLICATION NUMBER: PCT/EP01/02955
PRIORITY FILING DATE: 2001-03-15
PRIORITY APPLICATION NUMBER: DE 10013847.0
PRIORITY FILING DATE: 2000-03-15
PRIORITY APPLICATION NUMBER: DE 10019058.8
PRIORITY FILING DATE: 2000-04-06
PRIORITY APPLICATION NUMBER: DE 10019173.8
PRIORITY FILING DATE: 2000-04-07
PRIORITY APPLICATION NUMBER: DE 10032529.7
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: DE 10043826.1
PRIORITY FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 540
SEQ ID NO 170
LENGTH: 8245
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: 5109
OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-170
Query Match 0.64; Score 21; DB 12; Length 8245;

```

```

QY      3141  TTAGTATTGTTGTTTGTG 3161
Db      6632  TTAGTATTGTTGTTTGTG 6652

RESULT 7
US-10-001-843-87
; Sequence 87, Application US/10001843
; Publication No. US20020132255A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafterkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
; FILE REFERENCE: DDX-0267
; CURRENT APPLICATION NUMBER: US/10/001,843
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,992
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 87
;; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-001-843-87

```

Query Match	0.6%	Score 20;	DB 13;	Length 476;
Best Local Similarity	100.0%	Pred. No. 31;		
Matches 20; Conservative	0;	Mismatches	0;	Indels 0;
		Gaps	0;	

```

RESULT 8
US-09-736-734-1/c
; Sequence 1, Application US/09736734
; Patent No. US0020015983A1
; GENERAL INFORMATION:
; APPLICANT: ROMEO, Tony
; TITLE OF INVENTION: ESCHERICHIA COLI CSRB GENE, RNA ENCODED
; TITLE OF INVENTION: THEREBY, AND METHODS OF USE THEREOF
; FILE REFERENCE: 316082002001
; CURRENT APPLICATION NUMBER: US/09/736,734
; CURRENT FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: 09/112,584
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 60/052,372
; PRIOR FILING DATE: 1997-07-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-09-736-734-1

```

```

RESULT 9
US-10-396-911-1/c
: Sequence 1, Application US/10396911
: Publication No. US20030190752A1
: GENERAL INFORMATION:
: APPLICANT: KOMEO, TONY
: TITLE OF INVENTION: ESCHERICHIA COLI CSRB GENE, RNA ENCODED
: TITLE OF INVENTION: THEREAF, AND METHODS OF USE THEREOF
: FILE REFERENCE: 316082002001
: CURRENT APPLICATION NUMBER: US/10/396,911
: PRIOR FILING DATE: 2003-03-24
: PRIOR APPLICATION NUMBER: US/09/736,734
: PRIOR FILING DATE: 2000-12-13
: PRIOR APPLICATION NUMBER: 09/112,584
: PRIOR FILING DATE: 1998-07-09
: PRIOR APPLICATION NUMBER: 60/052,372
: PRIOR FILING DATE: 1997-07-11
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 654
: TYPE: DNA
: ORGANISM: Escherichia coli
US-10-396-911-1

```

Best Local Similarity 100.0%; Pred. No. 32;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3164 AAAATTGATGAAAGTTA 3183  
DB 153 AAAATTGATGAAAGTTA 134

RESULT 10  
US-10-074-475-133  
; Sequence 133, Application US/10074475  
; Publication No. US20030092898A1

GENERAL INFORMATION:  
APPLICANT: Salceda, Susana  
APPLICANT: Macina, Roberto  
APPLICANT: Hu, Ping  
APPLICANT: Reipon, Hervé  
APPLICANT: Kaita, Kalpana  
APPLICANT: Cafierkey, Robert  
APPLICANT: Sun, Yongming  
APPLICANT: Liu, Chenghua  
TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific  
TITLE OF INVENTION: Genes and Proteins  
FILE REFERENCE: DEX-0313  
CURRENT APPLICATION NUMBER: US/10/074,475  
CURRENT FILING DATE: 2002-02-13  
PRIOR APPLICATION NUMBER: 60/268,292  
PRIOR FILING DATE: 2001-02-13  
NUMBER OF SEQ ID NOS: 295  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 133  
LENGTH: 1023  
TYPE: DNA  
ORGANISM: Homo sapien  
US-10-074-475-133

Query Match 0.6%; Score 20; DB 14; Length 1023;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3136 TTTCCTTAGATTTTGTGTTT 3155  
DB 832 TTTCCTTAGATTTTGTGTTT 851

RESULT 11  
US-10-424-599-91177  
; Sequence 91177, Application US/10424599  
; Publication No. US20040031072A1

GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21 (53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 91177  
LENGTH: 1377  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53343C.1  
US-10-424-599-91177

Query Match 0.6%; Score 20; DB 12; Length 1377;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1368 CAGTGCATTGTCCAGCTTG 1387

DB 359 CAGTGCATTGTCCAGCTTG 378

RESULT 12  
US-10-282-122A-10949/C  
; Sequence 10949, Application US/10282122A  
; Publication No. US20040029129A1

GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Olsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 10949  
LENGTH: 1461  
TYPE: DNA  
ORGANISM: Borrelia burgdorferi  
US-10-282-122A-10949

Query Match 0.6%; Score 20; DB 12; Length 1461;  
Best Local Similarity 100.0%; Pred. No. 35;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2047 ATATATGCTTTAGTTTGT 2066  
DB 1073 ATATATGCTTTAGTTTGT 1054

RESULT 13  
US-10-424-599-91176  
; Sequence 91176, Application US/10424599  
; Publication No. US20040031072A1

GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53223)B  
 CURRENT APPLICATION NUMBER: US/10/424,599  
 CURRENT FILING DATE: 2003-04-28  
 NUMBER OF SEQ ID NOS: 285684  
 SEQ ID NO 91176  
 LENGTH: 1589  
 TYPE: DNA  
 ORGANISM: Glycine max  
 FEATURE:  
 OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53342C.1  
 US-10-424-599-91176

Query Match 0.6%; Score 20; DB 12; Length 1589;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1368 CAGTGCATTGTCAGCTTG 1387  
 DB 359 CAGTGCATTGTCAGCTTG 378

RESULT 14  
 US-10-108-260A-1756  
 Sequence 1756, Application US/10108260A  
 Publication No. US20040005560A1  
 GENERAL INFORMATION:  
 APPLICANT: HELIX RESEARCH INSTITUTE  
 TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA  
 FILE REFERENCE: H1-A0106  
 CURRENT APPLICATION NUMBER: US/10/108,260A  
 CURRENT FILING DATE: 2002-03-27  
 NUMBER OF SEQ ID NOS: 5458  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 1756  
 LENGTH: 2418  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-10-108-260A-1756

Query Match 0.6%; Score 20; DB 15; Length 2418;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3165 AAATTCATATGAAAGTAG 3184  
 DB 1988 AAATTCATATGAAAGTAG 2007

RESULT 15  
 US-09-947-953-1/c  
 Sequence 1, Application US/09947953  
 Patent No. US20020155101A1  
 GENERAL INFORMATION:  
 APPLICANT: DONAHUE, J. KEVIN  
 TITLE OF INVENTION: CARDIAC ARRHYTHMIA TREATMENT METHODS  
 FILE REFERENCE: 71699/56415  
 CURRENT APPLICATION NUMBER: US/09/947,953  
 CURRENT FILING DATE: 2001-09-06  
 PRIOR APPLICATION NUMBER: 60/230,311  
 PRIOR FILING DATE: 2001-09-06  
 PRIOR APPLICATION NUMBER: 60/235,889  
 PRIOR FILING DATE: 2001-06-05  
 NUMBER OF SEQ ID NOS: 2  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 1  
 LENGTH: 2985  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-947-953-1

Query Match 0.6%; Score 20; DB 9; Length 2985;

Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 CGAAAAAGGAAAAAAGG 90  
 DB 25 CGAAAAAGGAAAAAAGG 6

Search completed: April 9, 2004, 11:58:46  
 Job time : 929.457 secs



**This Page Blank (uspto)**



JOURNAL Patent: WO 0105951-A 2 25-JAN-2001;  
 AVENTIS CROPS SCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE  
 AGRONOMIQUE (FR)

FEATURES  
 source Location/Qualifiers  
 1. .1878  
 /organism="Arabidopsis thaliana"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:3702"

CDS

1. .1878  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC32419.1"  
 /db\_xref="GI:13158381"  
 /db\_xref="EMBL:CA032419"  
 /translation="MSSRAGPMSKKNVQGYRVEVQLVGLAETRLASSODDGGEM  
 EVISKKNKRGNTSGKTWVSNPNPAMGQOQGRSNVSGRNNVSGRNGRG  
 IONISGRGRLARKYDNNFVAPPVSRPPLGGNWMQAGSAGHTAVQEPVEDD  
 VDNASBERNDSDALDSDDDLDADYDSDVOSKHSRKNKMFKFGSLDLSIEQ  
 INEPRQWHPACONCPGATDWNHPLAHLARTGARVYLKHLRELAELEKDLQMRG  
 ASVLPCEIYGQWKGEDEKDEYIWPMPVYIEMARLRLAEMGLDRILAMQKRS  
 DKYEALRARSYGPQGHRSVLMFSSATGILEARLRELAEMGLDRILAMQKRS  
 FSGVROLYGFATKQDLDFNQSQKTRLEKLSYQEMVYKELROI SEMOOLNY  
 FSKNLSKONKHAUVLESLIMSEKLRTEENRIVROKQHEONREMAHREF  
 MDSIKOHERDRAKEENFEMLOOERAKVVGQOQONINPSNDCKRAEVSSTIER  
 QEKMEHFVERHEMLIKDEKQEDMKRHEEIPDLKEFDEALPOLMYKXGLHNE  
 D"

ORIGIN

Query Match 100.0%; Score 1878; DB 6; Length 1878;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAGTTCTAGGGCTGCTCCATGCTTAAGAAAAGAACTTCAGGGTGTATAGCCCT 60  
 1 ATGAGTTCTAGGGCTGCTCCATGCTTAAGAAAAGAACTTCAGGGTGTATAGCCCT 60  
 61 GAGGTGAACAGTTGGTTCAGAGTTTGGCAGGACGAGACTGGCTTCTTCAAGATGAT 120  
 61 GAGGTGAACAGTTGGTTCAGAGTTTGGCAGGACGAGACTGGCTTCTTCAAGATGAT 120  
 121 GAGAGAGAGTGGAGGCTCATTTTCCAAAGAAACAAGAACAAACGAGAAACATTTCTGA 180  
 121 GAGAGAGAGTGGAGGCTCATTTTCCAAAGAAACAAGAACAAACGAGAAACATTTCTGA 180  
 122 GAGAGAGAGTGGAGGCTCATTTTCCAAAGAAACAAGAACAAACGAGAAACATTTCTGA 180  
 181 AAAAATTGGGTTCTCAGAAATCGAATCTCTAGAGCTTGGGTGTCTCAGCAGAAAGG 240  
 181 AAAAATTGGGTTCTCAGAAATCGAATCTCTAGAGCTTGGGTGTCTCAGCAGAAAGG 240  
 181 AAAAATTGGGTTCTCAGAAATCGAATCTCTAGAGCTTGGGTGTCTCAGCAGAAAGG 240  
 241 AGAGGTAGCAACGTATCTGGGAGAGAAACAATGATTCGGGAGAGGTAAACGGCAATG 300  
 241 AGAGGTAGCAACGTATCTGGGAGAGAAACAATGATTCGGGAGAGGTAAACGGCAATG 300  
 241 AGAGGTAGCAACGTATCTGGGAGAGAAACAATGATTCGGGAGAGGTAAACGGCAATG 300  
 301 CGGGGCACTTCAAGGTACATATCTGCTCGGGGACGAGCGTTGAGCAGAAAGATATGATAC 360  
 301 CGGGGCACTTCAAGGTACATATCTGCTCGGGGACGAGCGTTGAGCAGAAAGATATGATAC 360  
 301 CGGGGCACTTCAAGGTACATATCTGCTCGGGGACGAGCGTTGAGCAGAAAGATATGATAC 360  
 361 AACTTTGTGGCAACCCCACTGTATCTGCGCCCTCTTTGAAAGAGAGATGAAATGGCAG 420  
 361 AACTTTGTGGCAACCCCACTGTATCTGCGCCCTCTTTGAAAGAGAGATGAAATGGCAG 420  
 361 AACTTTGTGGCAACCCCACTGTATCTGCGCCCTCTTTGAAAGAGAGATGAAATGGCAG 420  
 421 GCAAGAGAGAGTCTGCTCAGACACAGCTGTGAGAGATTTCTGAAGTGTGATATGAT 480  
 421 GCAAGAGAGAGTCTGCTCAGACACAGCTGTGAGAGATTTCTGAAGTGTGATATGAT 480  
 421 GCAAGAGAGAGTCTGCTCAGACACAGCTGTGAGAGATTTCTGAAGTGTGATATGAT 480  
 481 GTGATATATGCTTCTGAGAGAAAGATGATTCGATGCTTTGATGATATGATGAGCAG 540  
 481 GTGATATATGCTTCTGAGAGAAAGATGATTCGATGCTTTGATGATATGATGAGCAG 540  
 481 GTGATATATGCTTCTGAGAGAAAGATGATTCGATGCTTTGATGATATGATGAGCAG 540  
 541 CTTCGAAAGATGATTAAGTCTCGAGATGAGTCAAAAGAGCCATGATACGAAAGCAG 600  
 541 CTTCGAAAGATGATTAAGTCTCGAGATGAGTCAAAAGAGCCATGATACGAAAGCAG 600  
 541 CTTCGAAAGATGATTAAGTCTCGAGATGAGTCAAAAGAGCCATGATACGAAAGCAG 600  
 601 AATAAGTGGTCAAAAAGTTCTTTGGCAGCTTGATGATCTTGTGATCGACAGATTAAT 660

601 AATAAGTGGTCAAAAAGTTCTTTGGCAGCTTGATGATCTTGTGATCGACAGATTAAT 660  
 661 GAACCAAGAGAGAGTGGCATTTTCCAGCTTTGCAAGACGAGCTGGTGCATCGATTTG 720  
 661 GAACCAAGAGAGAGTGGCATTTTCCAGCTTTGCAAGACGAGCTGGTGCATCGATTTG 720  
 661 GAACCAAGAGAGAGTGGCATTTTCCAGCTTTGCAAGACGAGCTGGTGCATCGATTTG 720  
 721 TATTAACCTGCACCTCTTACTAGTCAATGCGAGAGCAAAAGAGCTAGCGAGTTAGCTC 780  
 721 TATTAACCTGCACCTCTTACTAGTCAATGCGAGAGCAAAAGAGCTAGCGAGTTAGCTC 780  
 721 TATTAACCTGCACCTCTTACTAGTCAATGCGAGAGCAAAAGAGCTAGCGAGTTAGCTC 780  
 721 TATTAACCTGCACCTCTTACTAGTCAATGCGAGAGCAAAAGAGCTAGCGAGTTAGCTC 780  
 781 CATTAAGAAATGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGTCAAT 840  
 781 CATTAAGAAATGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGTCAAT 840  
 781 CATTAAGAAATGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGTCAAT 840  
 841 CCTTGTGTGAGATTTATGCGAGAGTGAAGAGGTTTGGGTGAGAGTGAAGAAAGATTATGA 900  
 841 CCTTGTGTGAGATTTATGCGAGAGTGAAGAGGTTTGGGTGAGAGTGAAGAAAGATTATGA 900  
 841 CCTTGTGTGAGATTTATGCGAGAGTGAAGAGGTTTGGGTGAGAGTGAAGAAAGATTATGA 900  
 841 CCTTGTGTGAGATTTATGCGAGAGTGAAGAGGTTTGGGTGAGAGTGAAGAAAGATTATGA 900  
 901 ATTGTCTGGCTCCATTTGCTCATCATGATTAATTAATTAATTAATTAATTAATTAATTA 960  
 901 ATTGTCTGGCTCCATTTGCTCATCATGATTAATTAATTAATTAATTAATTAATTAATTA 960  
 901 ATTGTCTGGCTCCATTTGCTCATCATGATTAATTAATTAATTAATTAATTAATTAATTA 960  
 961 AAGTGGCTGGCAATGGGCAACAAAGCTGCTGAGATTAATTTGAGAACATGAGCTCTT 1020  
 961 AAGTGGCTGGCAATGGGCAACAAAGCTGCTGAGATTAATTTGAGAACATGAGCTCTT 1020  
 961 AAGTGGCTGGCAATGGGCAACAAAGCTGCTGAGATTAATTTGAGAACATGAGCTCTT 1020  
 1021 AAGAGAGAGGCAATCTTATGCTTCCAGAGGCTCATGAGAGTGAAGTGTCTGATGTTGAG 1080  
 1021 AAGAGAGAGGCAATCTTATGCTTCCAGAGGCTCATGAGAGTGAAGTGTCTGATGTTGAG 1080  
 1021 AAGAGAGAGGCAATCTTATGCTTCCAGAGGCTCATGAGAGTGAAGTGTCTGATGTTGAG 1080  
 1081 AGCAGTGCATCTGCTATTTTGGAGCCGAACGCTTCCACGGGAGTTAGCTGATGAGGAG 1140  
 1081 AGCAGTGCATCTGCTATTTTGGAGCCGAACGCTTCCACGGGAGTTAGCTGATGAGGAG 1140  
 1081 AGCAGTGCATCTGCTATTTTGGAGCCGAACGCTTCCACGGGAGTTAGCTGATGAGGAG 1140  
 1141 TTAAGTAAATTTGCTGAGGCTCAGAGGAGTATGTTTCTGAGAGTGTTCGCAACTG 1200  
 1141 TTAAGTAAATTTGCTGAGGCTCAGAGGAGTATGTTTCTGAGAGTGTTCGCAACTG 1200  
 1141 TTAAGTAAATTTGCTGAGGCTCAGAGGAGTATGTTTCTGAGAGTGTTCGCAACTG 1200  
 1201 TATGCTTCTTCAACAGAGAGATCTGAGCAATTAATTAATTAATTAATTAATTAATTA 1260  
 1201 TATGCTTCTTCAACAGAGAGATCTGAGCAATTAATTAATTAATTAATTAATTAATTA 1260  
 1201 TATGCTTCTTCAACAGAGAGATCTGAGCAATTAATTAATTAATTAATTAATTAATTA 1260  
 1261 ACAAGGCTGAATTTGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320  
 1261 ACAAGGCTGAATTTGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320  
 1261 ACAAGGCTGAATTTGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1320  
 1321 ATCTCTGAGAGCAATCAGAGCTGAACTAATTAAGAAACAAGCTTCAAAACAGAACAG 1380  
 1321 ATCTCTGAGAGCAATCAGAGCTGAACTAATTAAGAAACAAGCTTCAAAACAGAACAG 1380  
 1321 ATCTCTGAGAGCAATCAGAGCTGAACTAATTAAGAAACAAGCTTCAAAACAGAACAG 1380  
 1381 CACGCCAAGGCTTGAAGAAATCTCTGAAATTAATTAATTAATTAATTAATTAATTA 1440  
 1381 CACGCCAAGGCTTGAAGAAATCTCTGAAATTAATTAATTAATTAATTAATTAATTA 1440  
 1381 CACGCCAAGGCTTGAAGAAATCTCTGAAATTAATTAATTAATTAATTAATTAATTA 1440  
 1381 CACGCCAAGGCTTGAAGAAATCTCTGAAATTAATTAATTAATTAATTAATTAATTA 1440  
 1441 GAGGATTAATGATCTGAGAGAGAGAACTTAAGATCAGACATGAAGAGAGAG 1500  
 1441 GAGGATTAATGATCTGAGAGAGAGAACTTAAGATCAGACATGAAGAGAGAGAG 1500  
 1441 GAGGATTAATGATCTGAGAGAGAGAACTTAAGATCAGACATGAAGAGAGAGAG 1500  
 1501 ATGATGATCAGACAGAGTGTTCATGATTAATTAATTAATTAATTAATTAATTAATTA 1560  
 1501 ATGATGATCAGACAGAGTGTTCATGATTAATTAATTAATTAATTAATTAATTAATTA 1560  
 1501 ATGATGATCAGACAGAGTGTTCATGATTAATTAATTAATTAATTAATTAATTAATTA 1560  
 1561 GCAAGAGAGAGAAATTTGAGATGTTGAGCAGACAGAAAGTGCAGAGTGTGTCAG 1620  
 1561 GCAAGAGAGAGAAATTTGAGATGTTGAGCAGACAGAAAGTGCAGAGTGTGTCAG 1620  
 1561 GCAAGAGAGAGAAATTTGAGATGTTGAGCAGACAGAAAGTGCAGAGTGTGTCAG 1620  
 1621 CAGAGCAGAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1680  
 1621 CAGAGCAGAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1680  
 1621 CAGAGCAGAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1680  
 1681 TCAAGCTTCAAGATTTCAAGAGAGAGATGAGAGTGTGAGAGAGAGAGATG 1740

Db 1681 TCAAGCTTCATCGAGTTTCAAGAGAAAGAGATGAGAGAGCTTTGTGGAAGAGGAGATG 1740  
 Oy 1741 CTGATTAAGATCAAGAGAAAGATGAGATGAAAGAGAGGATCAAGAGAGATA 1800  
 Db 1741 CTGATTAAGATCAAGAGAAAGATGAGATGAAAGAGAGGATCAAGAGAGATA 1800  
 Oy 1801 TTGTATCTGGAGAAAGATTGATGAGGCTTTGGAACAGCTCATGTACAGATGGCTT 1860  
 Db 1801 TTGTATCTGGAGAAAGATTGATGAGGCTTTGGAACAGCTCATGTACAGATGGCTT 1860  
 Oy 1861 CACATGAGATGATGCA 1878  
 Db 1861 CACATGAGATGATGCA 1878

RESULT 2  
 BT004380 1909 bp mRNA linear PLN 14-FEB-2003  
 LOCUS Arabidopsis thaliana clone U20243 unknown protein (At5g23570) mRNA,  
 DEFINITION complete cds.  
 ACCESSION BT004380  
 VERSION BT004380.1 GI:28393932  
 KEYWORDS FLI CDNA.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE  
 AUTHORS 1 (bases 1 to 1909)  
 Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
 Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,  
 Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,  
 Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shimu,P.,  
 Southwick,A., Tripp,M.G., Wu,T., Shinzaki,K., Davis,R.W.,  
 Ecker,J.R. and Theologis,A.  
 Arabidopsis Open Reading Frame (ORF) Clones  
 Unpublished  
 2 (bases 1 to 1909)  
 Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
 Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,  
 Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,  
 Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shimu,P.,  
 Southwick,A., Tripp,M.G., Wu,T., Shinzaki,K., Davis,R.W.,  
 Ecker,J.R. and Theologis,A.  
 Direct Submission  
 Submitted (14-FEB-2003) Plant Gene Expression Center, 800 Buchanan  
 Street, Albany, CA 94710, USA

TITLE  
 JOURNAL The RIKEN Genomic Sciences Center (GSC) members carried out the  
 collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN  
 Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J.,  
 Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,  
 Hayashizaki,Y. and Shinzaki,K.

COMMENT  
 The Salk, Stanford, RGC (SSP) Consortium members constructed and  
 sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada,K.,  
 Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,  
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,  
 Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,  
 Palm,C.J., Shimu,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,  
 Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/RGC) and Seki,M. (RIKEN GSC) contributed equally  
 to this work. Shinzaki,K. (RIKEN GSC) and Theologis,A. (SSP  
 /RGC) contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome  
 submitted to Genbank.

FEATURES  
 source location/Qualifiers  
 1..1909  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"

/db xref="taxon:3702"  
 /chromosome="5"  
 /clone="U20243"  
 /ecotype="Columbia"  
 /note="This clone is in pUNI 51."  
 1..1909  
 /gene="At5g23570"  
 1..1878  
 /gene="At5g23570"  
 /codon\_start=1  
 /evidence=experimental  
 /product="unknown protein"  
 /protein\_id="AA042374.1"  
 /db xref="GI:28393932"  
 /translation="MSSRAGPMSKERNVGGYRPREVQLVGLAGTRLLASODDGEW  
 EVISKNNKPKPTSGKTMVSONSNPPRAKGGQGGKSGNSVGRNGVGRNGRG  
 IONATSGRRALSRKTDNNFVAPPVSRPPELGGKMWQKRGSAAGHTVAQEPDVEDD  
 VNASEENDSDALDDSDIDYDSVSSQSHSGSRKQNFKKFQSLDLSLEQ  
 INEPORWHCPACONGPGLIDWYNIHPLAHRATGARVKHRELAELERDLQWRC  
 ASVIPCSEIYQGMKGLSEDEKDYELIWPMPVYIINTRLDKNDKMLGNGQLELF  
 DKYELRARHSYGPQGHGMSVLMFESSATGYLEARHRELAEMGLRIANGQKSM  
 PEGGVRQLYGLATKODLDIPNHSQGRTRLKPELKSQVEMVYKELRQISENQOLN  
 FKQKSKQNHAKVLEBSLEINSEKLRTAEDNRIYRQRTKQHEONREEMADRF  
 MDSIKQIHERRAKSENEFMILQOQERAKVQOQONINPSSNDCKRAREVSSFLER  
 QEKEMEEFVEEREMLIKDQEKEMEDMKRHHEIFDLKEPFDEALEQLMYKRLHNE  
 D"  
 1879..1909  
 /gene="At5g23570"

## ORIGIN

3'UTR

Query Match 100.0%; Score 1878; DB 8; Length 1909;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ATGAGTTCTAGGCGCTGCTCCAAATGCTTAAGAAAGAACCTTCAGGCTGTTATAGCCT 60  
 Db 1 ATGAGTTCTAGGCGCTGCTCCAAATGCTTAAGAAAGAACCTTCAGGCTGTTATAGCCT 60  
 Oy 61 GAGGTGAACACTGGTTCAGGTTTGGAGGAGAGAGCTGGCTTCCAGAGATGAT 120  
 Db 61 GAGGTGAACACTGGTTCAGGTTTGGAGGAGAGAGCTGGCTTCCAGAGATGAT 120  
 Oy 121 GGAGGAGTGGGAGGCTATTTCCAGAAAGAACAAACAGAAACACTTCTGGA 180  
 Db 121 GGAGGAGTGGGAGGCTATTTCCAGAAAGAACAAACAGAAACACTTCTGGA 180  
 Oy 121 GAGAGAGATGGGAGGCTATTTCCAGAAAGAACAAACAGAAACACTTCTGGA 180  
 Db 121 GAGAGAGATGGGAGGCTATTTCCAGAAAGAACAAACAGAAACACTTCTGGA 180  
 Oy 181 AAAACTGGGTTTCTCAGAAATTCGATCTCTTGAAGCTTGGGCTGTCAGCAGAGG 240  
 Db 181 AAAACTGGGTTTCTCAGAAATTCGATCTCTTGAAGCTTGGGCTGTCAGCAGAGG 240  
 Oy 241 AAGAGTGAAGATCTGAGGAGAGAGAACATGTATCCGGAGAGTAAACGGCATGCT 300  
 Db 241 AAGAGTGAAGATCTGAGGAGAGAGAACATGTATCCGGAGAGTAAACGGCATGCT 300  
 Oy 301 CGGGGCAATTCAGCTAATATCTGCTCGGGACGAGCGTTGAGCAGAAATGATTAAC 360  
 Db 301 CGGGGCAATTCAGCTAATATCTGCTCGGGACGAGCGTTGAGCAGAAATGATTAAC 360  
 Oy 361 AACTTTGGGACCCCGACCTGTATCTCGCTCTTTGAAAGAGATGAAATTGGCAG 420  
 Db 361 AACTTTGGGACCCCGACCTGTATCTCGCTCTTTGAAAGAGATGAAATTGGCAG 420  
 Oy 421 GCAAGAGAGGTTCTGCTAGACACAGCTGACAGAGATTTCTGACGTGAGAGATGAT 480  
 Db 421 GCAAGAGAGGTTCTGCTAGACACAGCTGACAGAGATTTCTGACGTGAGAGATGAT 480  
 Oy 481 GTGATTAATCTTCTGAGAGAGAGATGATTCGATGCTTTGAGATGATTCGATGACGAC 540  
 Db 481 GTGATTAATCTTCTGAGAGAGAGATGATTCGATGCTTTGAGATGATTCGATGACGAC 540  
 Oy 541 CTTGCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
 Db 541 CTTGCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

QY	601	AATPAGTGGTTCAAAGAGTTCTTTGGCAGCTTGATGATGCTTGTGCATCGACGATTAAT	660
Db	601	AATPAGTGGTTCAAAGAGTTCTTTGGCAGCTTGATGATGCTTGTGCATCGACGATTAAT	660
QY	661	GAACCAACAGGCGAGTGGCATTTGTCCAGCTTGTCAAGAAACGAGCCGCGGCATGATTTGG	720
Db	661	GAACCAACAGGCGAGTGGCATTTGTCCAGCTTGTCAAGAAACGAGCCGCGGCATGATTTGG	720
QY	721	TATPAGCTGACCCCTCTACTAGCTCATGCGAGACAAAAGAGCTTAGCGATTACCTC	780
Db	721	TATPAGCTGACCCCTCTACTAGCTCATGCGAGACAAAAGAGCTTAGCGATTACCTC	780
QY	781	CATAGAGAAATGGCTGAGTTTATGAAAAGAGATCTACATAGAGAGCGGCATCTGTCAATT	840
Db	781	CATAGAGAAATGGCTGAGTTTATGAAAAGAGATCTACATAGAGAGCGGCATCTGTCAATT	840
QY	841	CCTTGTGTGATGATTTATGCGCAGTGGAGGGTTTGGGTGAGATGAAGATTTATCAA	900
Db	841	CCTTGTGTGATGATTTATGCGCAGTGGAGGGTTTGGGTGAGATGAAGATTTATCAA	900
QY	901	ATTGTCTGGCTCCAAATGCTATCATATCAATGAATCTGATTAAGGACGATTAACAT	960
Db	901	ATTGTCTGGCTCCAAATGCTATCATATCAATGAATCTGATTAAGGACGATTAACAT	960
QY	961	AAGTGGCTCGGCATGCGCAACCAAGAGCTCTGGAAATCTTCGACAAATATGAGCTCTT	1020
Db	961	AAGTGGCTCGGCATGCGCAACCAAGAGCTCTGGAAATCTTCGACAAATATGAGCTCTT	1020
QY	1021	AGAGCAGCCATTCCTATGTATGTCCACAGGGCCATCTGTGGGATGAGTGTCTGAAATTTGAG	1080
Db	1021	AGAGCAGCCATTCCTATGTATGTCCACAGGGCCATCTGTGGGATGAGTGTCTGAAATTTGAG	1080
QY	1081	ACAGTGGCCACTGGTATTTTGAAGCCGAACGCCTCCACCGGAGTTAGCTGAGATGGG	1140
Db	1081	ACAGTGGCCACTGGTATTTTGAAGCCGAACGCCTCCACCGGAGTTAGCTGAGATGGG	1140
QY	1141	TTAGATAGAAATTCCTGGGGGTGAGAGCGCAGTATGTTTTCTGAGAGTGTCCGCAACTG	1200
Db	1141	TTAGATAGAAATTCCTGGGGGTGAGAGCGCAGTATGTTTTCTGAGAGTGTCCGCAACTG	1200
QY	1201	TATGCTTCTCTTGCAACGAGCAATCTGGACATTAATCAATCAACACTCTCAAGCCAA	1260
Db	1201	TATGCTTCTCTTGCAACGAGCAATCTGGACATTAATCAATCAACACTCTCAAGCCAA	1260
QY	1261	ACAAGCTGAATTCGATGTGAATCATACAGAGATGTTGTAAAGAGAGCTGAGCGAG	1320
Db	1261	ACAAGCTGAATTCGATGTGAATCATACAGAGATGTTGTAAAGAGAGCTGAGCGAG	1320
QY	1321	ATCTCTGAGGACATGACGCTGAATCTTTAAGAACAGACTCTCAAAACAGAACAG	1380
Db	1321	ATCTCTGAGGACATGACGCTGAATCTTTAAGAACAGACTCTCAAAACAGAACAG	1380
QY	1381	CAGGCCAGGTGCTTGAGAAATCTCGAAATTTATGAGCGAAGCTGCTGTAACAATGCA	1440
Db	1381	CAGGCCAGGTGCTTGAGAAATCTCGAAATTTATGAGCGAAGCTGCTGTAACAATGCA	1440
QY	1441	GAGGATTAATCGATCTGAGAGACGAAACTTAAGATGACAGCATGAACAGGAAAGAG	1500
Db	1441	GAGGATTAATCGATCTGAGAGACGAAACTTAAGATGACAGCATGAACAGGAAAGAG	1500
QY	1501	ATGATGTCACACGACAGGTTTTTCATGATATCAATCAACAGATCCATGAAAGAGAGAC	1560
Db	1501	ATGATGTCACACGACAGGTTTTTCATGATATCAATCAACAGATCCATGAAAGAGAGAC	1560
QY	1561	GCAAGAGAGAGAAATTTCCGATGAGATGTTGACAGCAGAGAACGTCGTAAGAACTGCA	1620
Db	1561	GCAAGAGAGAGAAATTTCCGATGAGATGTTGACAGCAGAGAACGTCGTAAGAACTGCA	1620
QY	1621	CAGACGACGAACATTAAATCCCTAGCAATGAGATTTGCCAAGAGAGACTGAGAGATG	1680
Db	1621	CAGACGACGAACATTAAATCCCTAGCAATGAGATTTGCCAAGAGAGACTGAGAGATG	1680

FEATURES	location/Qualifiers
source	1. .2162

Annotation based on July 2002 version of the Arabidopsis genome submitted to Genbank.

```
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAF14-93-K05 (R20243)"
/ecotype="Columbia"
/Note="This clone is in a modified pBluescript vector2
(lambda ps) as a BamHI/XhoI insert."
1..2162
/gene="At5g23570"
5'UTR
1..91
/gene="At5g23570"
CDS
92..1969
/gene="At5g23570"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="A022757.1"
/db_xref="GI:27754623"
/translation="MSBRAGPMSEKKNVGVPEVQLVGLAGTRIASDDGGM
EVSKKNKPKGNTSGKTWVSONSPRAAGGQOGRGSVSGRNNVSGRGNGR
IQANISGRGALSRKTDNPFVAPPSRPLEBGRNKAQAGSAHITAVEPFDVED
VDNASENDSDALDDSDDDASDDYSDVSQSHSRKQKPKFPGSLDLSIBQ
INBPQRMHCPAQONGGALDWNHPLAHARTKARVQIARELAEVLKDLQWRG
ASVTPGKIGYQWKGIGEDKDEYEVIPPVIVIMNTRLDKDNDKMLGMNOELLEY
DKYALRARSYVPGQHRKSVLMFESSATGVLEAEIRHLELAEMDLRIAMGQKSM
FGSVRLVGLATKODLDIFNHSQKTPKRELSYQEMVYKEIROI SEDNOQNTY
FKRLSKONKHAIVLESLEFIMSEKLRRTEDNRIYRQRTRMQHEQREMDHDF
MDSIKQIHERDAKEENFEMILQOQERAKVVGQOQINPSSNDCRRRAEVSTLEF
QEKEMEFVEREMLIKDQEKMDMKRHEBI FDLKEDBALQMLMYKHLNHD
D"
1970..2162
/gene="At5g23570"
misc_difference
2146
/gene="At5g23570"
/Note="not present in genomic sequence"

ORIGIN

Query Match      100.0%; Score 1878; DB 8; Length 2162;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTCTAGGCTGCTCCATGTCATAGGAAAAGAACGTTGAGGTGTTATAGGCT 60
DB 92 ATGAGTTCTAGGCTGCTCCATGTCATAGGAAAAGAACGTTGAGGTGTTATAGGCT 151
QY 61 GAGGTGAACAGTTGTTCAAGGTTTGAGGAGGACGAGACTGCTTCTTCAAGATGAT 120
DB 152 GAGGTGAACAGTTGTTCAAGGTTTGAGGAGGACGAGACTGCTTCTTCAAGATGAT 211
QY 121 GGAGGAGAGTGGAGGTCATTTCCAGAGAAACAAGAACCAAGAAACATTTCTGGA 180
DB 212 GGAGGAGAGTGGAGGTCATTTCCAGAGAAACAAGAACCAAGAAACATTTCTGGA 271
QY 181 AAAACTTGGGTTTCTCAGATTGCAATCTCTAGAGCTTGGGTTGTCAGACGAGG 240
DB 272 AAAACTTGGGTTTCTCAGATTGCAATCTCTAGAGCTTGGGTTGTCAGACGAGG 331
QY 241 AGAGTAGCAACGTATCTGGGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATGGT 300
DB 332 AGAGTAGCAACGTATCTGGGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATGGT 391
QY 301 CGGGGCAATTCACCTAATCATATCTGTCGGGAGAGAGCTTGAACGAAATGTATTAAC 360
DB 392 CGGGGCAATTCACCTAATCATATCTGTCGGGAGAGAGCTTGAACGAAATGTATTAAC 451
QY 361 AACCTTGGCACCCCACTGTATCTGCTCTCTTTGGAAGAGGATGAATTTGGCAG 420
DB 452 AACCTTGGCACCCCACTGTATCTGCTCTCTTTGGAAGAGGATGAATTTGGCAG 511
QY 421 GCAAGAGAGAGTTCTGTCAGCAACAGCTGTGAGAGATTTCTGACGTGAGAGATGAT 480
DB 512 GCAAGAGAGAGTTCTGTCAGCAACAGCTGTGAGAGATTTCTGACGTGAGAGATGAT 571
```

```
QY 481 GTGATTAATGCTTCTGAGGAAGAGATGATTCGATGCTTTGGATGATTTCTGATGACGAC 540
DB 572 GTGATTAATGCTTCTGAGGAAGAGATGATTCGATGCTTTGGATGATTTCTGATGACGAC 631
QY 541 CTTGCAAGTATGATTTATGACTCGGATGTGAGTCAAAAGAGCCATGATCACGAAAGCAG 600
DB 632 CTTGCAAGTATGATTTATGACTCGGATGTGAGTCAAAAGAGCCATGATCACGAAAGCAG 691
QY 601 AATTAATGCTTCAAAAAGTTCTTTGGCAGCTTGGATTAAGCTTGTCATCGACGATTAAT 660
DB 692 AATTAATGCTTCAAAAAGTTCTTTGGCAGCTTGGATTAAGCTTGTCATCGACGATTAAT 751
QY 661 GAACCAAGAGCAGTGGCATTTGTCAGGCTTTCAGAGAGAGCCTGTCATCGATTTG 720
DB 752 GAACCAAGAGCAGTGGCATTTGTCAGGCTTTCAGAGAGAGCCTGTCATCGATTTG 811
QY 721 TATTAACCTGACCCCTTACTAGCTCATGCGAGGACAAAAGAGCTTAGGCGATTAAAGCTC 780
DB 812 TATTAACCTGACCCCTTACTAGCTCATGCGAGGACAAAAGAGCTTAGGCGATTAAAGCTC 871
QY 781 CATAGGAATTTGCTTAAGTTTAAAGAGATCTACATGAGAGCGCATCTGTCATT 840
DB 872 CATAGGAATTTGCTTAAGTTTAAAGAGATCTACATGAGAGCGCATCTGTCATT 931
QY 841 CCTTGTGTGAGATTATAGGGCAGTGGAAAGGTTTGGTGAAGATGAAGAAAGATTATGAA 900
DB 932 CCTTGTGTGAGATTATAGGGCAGTGGAAAGGTTTGGTGAAGATGAAGAAAGATTATGAA 991
QY 901 ATTGTGCTGCTTCAATGTCATCATGAAATCTAGACTGATTAAGACGATTAACGAT 960
DB 992 ATTGTGCTGCTTCAATGTCATCATGAAATCTAGACTGATTAAGACGATTAACGAT 1051
QY 961 AATGCTGCTGCGATGCGGCAACCAAGCTGCTGGAATATCTTGACCAAGTATGAGCTCTT 1020
DB 1052 AATGCTGCTGCGATGCGGCAACCAAGCTGCTGGAATATCTTGACCAAGTATGAGCTCTT 1111
QY 1021 AGAGCAAGCATTCTCATAGTGCACAGAGGCAATCGGAGTGAAGTTCGATGTTTGAAG 1080
DB 1112 AGAGCAAGCATTCTCATAGTGCACAGAGGCAATCGGAGTGAAGTTCGATGTTTGAAG 1171
QY 1081 AGCAGTGCACCTGCTATTTTGAAGGCGGACGCTTCCACCGGAGATTAGCTGATGAGG 1140
DB 1172 AGCAGTGCACCTGCTATTTTGAAGGCGGACGCTTCCACCGGAGATTAGCTGATGAGG 1231
QY 1141 TTGATTAAGATTGCTGCGGCTCAAGAGCGAGTATTTTCTGAGAGTGTGCGCACTG 1200
DB 1232 TTGATTAAGATTGCTGCGGCTCAAGAGCGAGTATTTTCTGAGAGTGTGCGCACTG 1291
QY 1201 TATGCTTCTTGAACGAGAGCAAGATCTGACATTAATCAACACTCTCAAGGCATA 1260
DB 1292 TATGCTTCTTGAACGAGAGCAAGATCTGACATTAATCAACACTCTCAAGGCATA 1351
QY 1261 ACAAGCTGAATTTGAGTTGAATCATACCAAGAGATGTTTAAAGAGCTGAGGCGAG 1320
DB 1352 ACAAGCTGAATTTGAGTTGAATCATACCAAGAGATGTTTAAAGAGCTGAGGCGAG 1411
QY 1321 ATCTTGAGGACAACTGACGCTGAACTTATTAAGAACAGCTCTCAAAACAGAACAG 1380
DB 1412 ATCTTGAGGACAACTGACGCTGAACTTATTAAGAACAGCTCTCAAAACAGAACAG 1471
QY 1381 CAGCGCAAGGCTTGAAGAACTCTGGAATTTATGAGGAGAGAGCTGCTGAACCTGA 1440
DB 1472 CAGCGCAAGGCTTGAAGAACTCTGGAATTTATGAGGAGAGAGCTGCTGAACCTGA 1531
QY 1441 GAGGATTAATCGATCGTGAAGACAGAACTTAAGATGACATGAACAGAACAGGAGAG 1500
DB 1532 GAGGATTAATCGATCGTGAAGACAGAACTTAAGATGACATGAACAGAACAGGAGAG 1591
QY 1501 ATGATGACACAGCAGGTTTTTTCATGATTAATCAACAGATCAATGAAGAGAGAC 1560
DB 1592 ATGATGACACAGCAGGTTTTTTCATGATTAATCAACAGATCAATGAAGAGAGAC 1651
```

QY 1561 GCMAAGAGAGAAATTTCCAGATGTTGACAGCAGAGAAAGTCCAAAGTGTGTTGGCCAG 1620  
 DB 1652 GCMAAGAGAGAAATTTCCAGATGTTGACAGCAGAGAAAGTCCAAAGTGTGTTGGCCAG 1711  
 QY 1621 CAGCAGAGAAATTTATCCCTCTTACCAATGACATGTCGCAAGAGAGCTGAGGAAGTG 1680  
 DB 1712 CAGCAGAGAAATTTATCCCTCTTACCAATGACATGTCGCAAGAGAGCTGAGGAAGTG 1771  
 QY 1681 TCAGAGCTTCATGAGTTTCAGAGAAAGAGATGAGAGAGATTTGTGTGAGAGAGAGAGAG 1740  
 DB 1772 TCAGAGCTTCATGAGTTTCAGAGAAAGAGATGAGAGAGATTTGTGTGAGAGAGAGAGAG 1831  
 QY 1741 CTTATTAAGAAATGAGAGAAAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAT 1800  
 DB 1832 CTTATTAAGAAATGAGAGAAAGAGATGAGAGAGATGAGAGAGATGAGAGAGATGAGAGAT 1891  
 QY 1801 TTTGATCTGAGAGAAAGATTTGATGAGAGCTTTGGAACAGCTCATGTGACAGATGAGCTTT 1860  
 DB 1892 TTTGATCTGAGAGAAAGATTTGATGAGAGCTTTGGAACAGCTCATGTGACAGATGAGCTTT 1951  
 QY 1861 CACATGAGAGATGATGCA 1878  
 DB 1952 CACATGAGAGATGATGCA 1969

RESULT 4  
 AF239719 2254 bp DNA linear PLN 02-JUN-2000  
 LOCUS Arabidopsis thaliana SGS3 gene, complete cds.  
 DEFINITION AF239719  
 VERSION AF239719.1 GI:8164029  
 KEYWORDS  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 2254)  
 Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C.,  
 Morel, J.B., Jonet, D., Lacombe, A.M., Nikic, S., Picault, N.,  
 Remoue, K., Santal, M., Vo, T.A. and Vaucheret, H.  
 Arabidopsis SGS2 and SGS3 genes are required for  
 posttranscriptional gene silencing and natural virus resistance  
 Cell 101 (5), 533-542 (2000)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 SOURCE

1. 2254  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /cultivar="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="v"  
 join(1..963,1037..1328,1440..1684,1787..1953,2044..2254)  
 /product="SGS3"  
 /codon\_start=1  
 /product="SGS3"  
 /protein\_id="AF23960.1"  
 /db\_xref="GI:8164030"  
 /translation="WSRAGPMSKKNVQGYRPEVBLVQGLAGTRLASODDGGEM  
 EVISKKNKPKNTGKTWVSNPNPAGGQOQGRSNVSGNNVSGRNGG  
 IOANISGRGALSRKYDNNFVAPVPVSPPLGGNNWARGSAQHTAVQFPEVED  
 VDNASEENDSDALDDSDLDASDDYDVSQSHSGKQKMFKEFGSLDSISIBQ  
 INEPORWHCAPCONGPAIDVYNLHPLLAHRTGARGVRLHLEVEKDIQMG  
 ASVTGCEIYQGMKGLGDEKDYETVMPMVTIMTTRLDKDNDKMGWQELLEY  
 DKTEALRAHSTGPGHKGMSVLMFESSATGTLEBRRLHRLAENGDLDRIMGQKRM  
 FSGVROLVYGLATKQDLIDFNQHSQGRILKFLKSLQEMVWVLELROIISDNOQNT

## ORIGIN

Query Match 51.3%; Score 963; DB 8; Length 2254;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FNKLKSNKRNKARVLEESLEIMSEKLRFLAEDNRIVORTKQMHQNEBMDAHNREF  
 MDSIKQIHERRDAKKEENFEMLOOQERAKRVQOQOINIBSSNDCKRAEVSFTLF  
 QEKMEBFVEREMLIKDQEKMEBDMKRRHHEIFILKXEPDEALBQLMYKGLHNE  
 D"  
 1 ATGAGTTTACGAGGCTGGTCCAAATGCTTAGGAAAGAAAGCTTCAGGNGTTATAGGCT 60  
 1 ATGAGTTTACGAGGCTGGTCCAAATGCTTAGGAAAGAAAGCTTCAGGNGTTATAGGCT 60  
 QY 61 GAGGTTGAACAGTTGTTCAAGTTTGGCAGGAGCAGACTGCTTTCTTCAAGATGAT 120  
 DB 61 GAGGTTGAACAGTTGTTCAAGTTTGGCAGGAGCAGACTGCTTTCTTCAAGATGAT 120  
 QY 121 GGAGAGAGTGGAGAGTCAATTTCCAGAGAAACAAGAACCAACAGAGAAACATTTTGA 180  
 DB 121 GGAGAGAGTGGAGAGTCAATTTCCAGAGAAACAAGAACCAACAGAGAAACATTTTGA 180  
 QY 181 AAAAATTTGGTTTCTCAAGATTTCAATCTCTGAGCTTGGGTTGTCAGCAGCAAGG 240  
 DB 181 AAAAATTTGGTTTCTCAAGATTTCAATCTCTGAGCTTGGGTTGTCAGCAGCAAGG 240  
 QY 241 AGAGTTAGCAACGTATCTGAGAGAGAAACAATGTATCCGAGAGAGTTACCGCAATGT 300  
 DB 241 AGAGTTAGCAACGTATCTGAGAGAGAAACAATGTATCCGAGAGAGTTACCGCAATGT 300  
 QY 301 CCGGAGATTCAAGCTAACATATCTGTTGGGAGCAGAGGTTGAGCAGAAATATGATAC 360  
 DB 301 CCGGAGATTCAAGCTAACATATCTGTTGGGAGCAGAGGTTGAGCAGAAATATGATAC 360  
 QY 361 AACTTTGGGACCCCACTGTATCTGACCTCTCTTGAAGAGATGAAATTCGAG 420  
 DB 361 AACTTTGGGACCCCACTGTATCTGACCTCTCTTGAAGAGATGAAATTCGAG 420  
 QY 421 GCMAAGAGAGTTCGCTCTGACACACAGCTGTGCAGAGTTCCTGACGTGAGATAT 480  
 DB 421 GCMAAGAGAGTTCGCTCTGACACACAGCTGTGCAGAGTTCCTGACGTGAGATAT 480  
 QY 481 GTGATATGCTTCTGAGAGAAAGAAATATTCGATTCGTTGAGATTCGATGAGCAG 540  
 DB 481 GTGATATGCTTCTGAGAGAAAGAAATATTCGATTCGTTGAGATTCGATGAGCAG 540  
 QY 541 CTTGCAAGTATGATTTATGATCTCGATGAGATCAAAAGAGCATGATCAAGAAAGCAG 600  
 DB 541 CTTGCAAGTATGATTTATGATCTCGATGAGATCAAAAGAGCATGATCAAGAAAGCAG 600  
 QY 601 AATAGTGTTCAAAAGTCTTTTGGCAGCTTGAATAGCTTGTGATGAGAGATTAAT 660  
 DB 601 AATAGTGTTCAAAAGTCTTTTGGCAGCTTGAATAGCTTGTGATGAGAGATTAAT 660  
 QY 661 GAACCAAGAGAGAGAGATGTCAGCTTTCAGAAAGCAGCTGTCATGATGAG 720  
 DB 661 GAACCAAGAGAGAGAGATGTCAGCTTTCAGAAAGCAGCTGTCATGATGAG 720  
 QY 721 TATACCTGACCCCTTACTAGCTCATGCGAGAGACAAAAGAGACTAGGAGTTAAGCTC 780  
 DB 721 TATACCTGACCCCTTACTAGCTCATGCGAGAGACAAAAGAGACTAGGAGTTAAGCTC 780  
 QY 781 CATAGAGATTTGGCTGAAGTTTAAAGAAAGATCTACATAGAGGCGCATCTGCAT 840  
 DB 781 CATAGAGATTTGGCTGAAGTTTAAAGAAAGATCTACATAGAGGCGCATCTGCAT 840  
 QY 841 CTTGAGAGAGATTTATGAGGAGTGAAGGTTTGGTGAAGATGAAGATTAATGAA 900  
 DB 841 CTTGAGAGAGATTTATGAGGAGTGAAGGTTTGGTGAAGATGAAGATTAATGAA 900  
 QY 901 ATTGTGAGCTCCAAATGTCATCATGAATATCTAGACTGATTAAGAGACGATTAAGAT 960  
 DB 901 ATTGTGAGCTCCAAATGTCATCATGAATATCTAGACTGATTAAGAGACGATTAAGAT 960



```

QY      961 AAG 963
      |||
Db      961 AAG 963

RESULT 5
AX078760      3275 bp      DNA      linear      PAT 22-FEB-2001
LOCUS      Sequence 1 from Patent WO0105951.
DEFINITION  AX078760
ACCESSION   AX078760.1 GI:13158379
VERSION     AX078760.1
KEYWORDS    '
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE   1
AUTHORS     Beclin,C., Elmayan,T. and Vaucheret,H.
TITLE       Novel sgs3 plant gene and use thereof
JOURNAL     Patent: WO 0105951 A 1 25-JAN-2001;
            AVENTIS CROPS/SCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
            AGRONOMIQUE (FR)
FEATURES
            location/Qualifiers
            source      1..3275
                        /organism="Arabidopsis thaliana"
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:3702"
                        693..715
                        /note="p356AD"
                        complement(2926..2952)
                        /note="p356Y"

ORIGIN
Query Match      51.3%; Score 963; DB 6; Length 3275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATGAGTTTATGAGGCTGTCCATGCTCAAGAAAAGAAAGCTTCAAGGTGTTATAGCCCT 60
Db      696 ATGAGTTTATGAGGCTGTCCATGCTCAAGAAAAGAAAGCTTCAAGGTGTTATAGCCCT 755

QY      61 GAGGTGAACAGTGGTTCAGAGTTTGGACGAGGACGAGTGGCTTCTTCAAGATGAT 120
Db      756 GAGGTGAACAGTGGTTCAGAGTTTGGACGAGGACGAGTGGCTTCTTCAAGATGAT 815

QY      121 GGAGGAGTGGAGGTTCATTTCCAAAGAACAGAACAAACAGGAAACACTTCTGGA 180
Db      816 GGAGGAGTGGAGGTTCATTTCCAAAGAACAGAACAAACAGGAAACACTTCTGGA 875

QY      181 AAAACTTGGGTTTCTCAGATTTGAATCTCTTCAAGCTTGGGTTGTCACGCAAGGG 240
Db      876 AAAACTTGGGTTTCTCAGATTTGAATCTCTTCAAGCTTGGGTTGTCACGCAAGGG 925

QY      241 AGAGTAGCAACGATCTGGGAGAGGAAACATGATCCGGGAGAGGTAAACGGCAATGCT 300
Db      936 AGAGTAGCAACGATCTGGGAGAGGAAACATGATCCGGGAGAGGTAAACGGCAATGCT 995

QY      301 CGGGGCAATTCAGCTAACATATCTGTCGGGAGACGAGCTTGAAGCAAAAGTATGATAC 360
Db      996 CGGGGCAATTCAGCTAACATATCTGTCGGGAGACGAGCTTGAAGCAAAAGTATGATAC 1055

QY      361 AACTTTTGGACACCCCACTGATATCTCGCCCTCTTTGAAGAGAGATGGAATTGGCAG 420
Db      1056 AACTTTTGGACACCCCACTGATATCTCGCCCTCTTTGAAGAGAGATGGAATTGGCAG 1115

QY      421 GCAAGAGAGAGTCTGTCTGACACACAGCTGTGAGAGAGTTTCTGACGTGAGAGATGAT 480
Db      1116 GCAAGAGAGAGTCTGTCTGACACACAGCTGTGAGAGAGTTTCTGACGTGAGAGATGAT 1175

QY      481 GTGGAATATGCTTCTGAGAGAGAGATGATTCGATGCTTTGATGATGATGATGACGAC 540
      |||

```

```

Db      1176 GTGGAATATGCTTCTGAGAGAGAGATGATTCGATGCTTTGAGATTCGATGACGAC 1235
QY      541 CTGCAAGTATGATTTATGATCTGAGTGTAGTCAAAAAGAGCCATGATCAGAAAGCAG 600
      |||
Db      1236 CTGCAAGTATGATTTATGATCTGAGTGTAGTCAAAAAGAGCCATGATCAGAAAGCAG 1295
QY      601 AATTAAGTGTCAAAAAGTCTTTGGCAGCTTGATAGCTTGTGATCGACGATTAAT 660
      |||
Db      1296 AATTAAGTGTCAAAAAGTCTTTGGCAGCTTGATAGCTTGTGATCGACGATTAAT 1355
QY      661 GAACCAAGAGGACGATGTCATGTCAGTCTTGCAGAACGACCTGTGCATCAATTGG 720
Db      1356 GAACCAAGAGGACGATGTCATGTCAGTCTTGCAGAACGACCTGTGCATCAATTGG 1415
QY      721 TATTAACCTGACCCCTTACTAGTCAATGAGAGACAAAAGAGCTGAGGAGTTAAGCTC 780
      |||
Db      1416 TATTAACCTGACCCCTTACTAGTCAATGAGAGACAAAAGAGCTGAGGAGTTAAGCTC 1475
QY      781 CATAGAAATTGGCTGAACTTTTGAAGAAAGATCTACAGATGAGAGCCATCTGTCAAT 840
      |||
Db      1476 CATAGAAATTGGCTGAACTTTTGAAGAAAGATCTACAGATGAGAGCCATCTGTCAAT 1535
QY      841 CTTGTGTGAGATTTATGAGGACGTGAAGAGGTTGGTGTGAGAGTAAAGATTATGAA 900
Db      1536 CTTGTGTGAGATTTATGAGGACGTGAAGAGGTTGGTGTGAGAGTAAAGATTATGAA 1595
QY      901 ATTGTGCGCTCCCAATGTCATCATGAAATCTAGACCTGATAGAGACGATTAAGCAT 960
      |||
Db      1596 ATTGTGCGCTCCCAATGTCATCATGAAATCTAGACCTGATAGAGACGATTAAGCAT 1655
QY      961 AAG 963
      |||
Db      1656 AAG 1658

RESULT 6
AB025633      81365 bp      DNA      linear      PLN 27-DEC-2000
LOCUS      Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MQM1.
DEFINITION  AB025633 BAO00015
ACCESSION   AB025633.2 GI:10178221
VERSION     AB025633.2
KEYWORDS    '
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE   1 (sites)
AUTHORS     Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H.
            and Tabata,S.
TITLE       Structural analysis of Arabidopsis thaliana chromosome 5. X.
            Sequence features of the regions of 3,076,755 bp covered by sixty
            P1 and TAC clones
JOURNAL     DNA Res. 7 (1), 31-63 (2000)
MEDLINE     20181125
PUBMED     10718197
REFERENCE   2 (bases 1 to 81365)
AUTHORS     Nakamura,Y.
TITLE       Direct Submission
            Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
            On Sep 15, 2000 this sequence version replaced gi:4589439.
COMMENT     Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/sgd_graph.cgi?c=MQM1
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers. Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/Grail-1.3/),

```





/codon\_start=1  
 /evidence=not experimental  
 /protein\_id="BAA97243.1"  
 /db\_xref="GI:8809702"  
 /translation="MAPAPAGCCTAAVAAAGADGAPQOQGRGSGTIGTIVAVYV  
 ASKPSPEKPKMDPTAPASQIMNTLPHKRSGLDMVYVLEEQKEDGSDGIVLVYHETV  
 IPYAVMTBRSITKSLITYTPSETLONNSGLYAHITFAASGPFIDTDBERYOPLNSFS  
 THAVATVPKQKKKKKSLUGSPKDSSESEPEVNGKSPSEVEVETISKLPY  
 VTIYLVDPFOADLLVEPTITVYPTIYEPHEFWLRKRKLPVNEVETPLNIEIS  
 ISMKWQKLEQVQDOSFQOMRGLSAGKSVLNFITQCFYPIYLNDNTSNMILLSGQVQ  
 DFLAKNIOLEFNNKRSKBSAGKSVLNFITQCFYPIYLNDNTSNMILLSGQVQ  
 IEFMYIGAMRIEVDNSGIMRIPLRPHDEBSYASNTKESYDIAIFELSYALLVIG  
 SYSLASVLRHKSWMYSLISLTSQVYMGFIMMCCQLFNTYGLKSAVHLIPROMTYKK  
 LKTIIDDEPAFYIKMPLILRHLSVFRDVIPILYIGRWVYVPVDKTRVNEFGCGEDED  
 AKKTLIDDEEEDKKTN"  
 cds complement(join(29493)..29576,29661..29685,29985..30268  
 30341..30493,30579..30722,30956..31484))  
 /note="gene\_id:MQM1.15

Query Match	51.3%	Score 963	DB 8	Length 81365	
Best Local Similarity	100.0%	Pred. No. 0			
Matches 963	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	1	ATGAGTTCTAGGGCTGGTCCAAATGTCTTAAGAAAAGAACGTTCAAGGTGGTTATAGAGCCT	60		
Db	22672	ATGAGTTCTAGGGCTGGTCCAAATGTCTTAAGAAAAGAACGTTCAAGGTGGTTATAGAGCCT	22733		
QY	61	GAGGTGAACAATTGGTTCAGAGTTTGGCAGGACGAGACTGGCTTCTTCAACAGATGAT	120		
Db	22732	GAGGTGAACAATTGGTTCAGAGTTTGGCAGGACGAGACTGGCTTCTTCAACAGATGAT	22791		
QY	121	CGAGGAGATGGGAGGTCATTTTCCAGAAAGAACAAAGAACAAACACAGAAACCTTTGGA	180		
Db	22792	GGAGGAGATGGGAGGTCATTTTCCAGAAAGAACAAAGAACAAACACAGAAACCTTTGGA	22851		
QY	181	AAAACCTGGGTTTCTCGAATTCGAATCCCTAGACCTGGGGTGGTCAAGCAGCAAGG	240		
Db	22852	AAAACCTGGGTTTCTCGAATTCGAATCCCTAGACCTGGGGTGGTCAAGCAGCAAGG	22911		
QY	241	AGAGTNGCAACGTATCTGGGAGAGGAACAATGTAATCCGGAGAGGTAAACGCAATGGT	300		
Db	22912	AGAGTNGCAACGTATCTGGGAGAGGAACAATGTAATCCGGAGAGGTAAACGCAATGGT	22971		
QY	301	CGGGGCAATTCAAGCTAACATATCTGGTCCGGGAGCAGACCTTGAGCGAAGAAAGTTGATAC	360		
Db	22972	CGGGGCAATTCAAGCTAACATATCTGGTCCGGGAGCAGACCTTGAGCGAAGAAAGTTGATAC	23033		
QY	361	AACCTTGTGGCACCCCACTGTATCTCGCCCTCTTTGGAGAGAGATGGAATTGGCAG	420		
Db	23032	AACCTTGTGGCACCCCACTGTATCTCGCCCTCTTTGGAGAGAGATGGAATTGGCAG	23091		
QY	421	GCAAGAGAGGTTCTGCTCAGCACACACAGCTGTGCAGAGATTTCTGACGTGAGATGAT	480		
Db	23092	GCAAGAGAGGTTCTGCTCAGCACACACAGCTGTGCAGAGATTTCTGACGTGAGATGAT	23151		
QY	481	GTGATATATGCTTCTGAGAAAGATGATTTCCGATGCTTTGGATGATTTCTGATGACAC	540		
Db	23152	GTGATATATGCTTCTGAGAAAGATGATTTCCGATGCTTTGGATGATTTCTGATGACAC	23211		
QY	541	CTTGCAGGTATGATTTATGACTGGATGTGACCAAAAGCCATGGAATCAACGAAAGCAG	600		
Db	23212	CTTGCAGGTATGATTTATGACTGGATGTGACCAAAAGCCATGGAATCAACGAAAGCAG	23271		
QY	601	AATTAAGGTTCAAAAAGTCTTTGGCAGCTTGGATAGCTTGTGCATCGACAGATTAAT	660		
Db	23272	AATTAAGGTTCAAAAAGTCTTTGGCAGCTTGGATAGCTTGTGCATCGACAGATTAAT	23333		
QY	661	GAAACACAGAGGCAATGGCAATTTGTCAAGCTTGTCAAGCGACCTGTGCAATGATGG	720		
Db	23332	GAAACACAGAGGCAATGGCAATTTGTCAAGCTTGTCAAGCGACCTGTGCAATGATGG	23391		
QY	721	TATTAACCTGCAACCTCTACTAGCTCATGCGAGCAAAAGAGCTTGGGAGTTAAGCTC	780		

Db	23392	TATTAACCTGCACCCCTTACTAGCTCATGCGAGACAAGAAAGAACTGAGCGAGTTAAGCTC	23451
Qy	781	CATAGAGAAATTGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGCATT	840
Db	23452	CATAGAAATTTGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGCATT	23511
Qy	841	CCTTGTGTTAGATTTATATGCGCAGTGGAAAGGCTTTGGTGAGGATGAAAAGATTTATGAA	900
Db	23512	CCTTGTGTTAGATTTATATGCGCAGTGGAAAGGCGTTTGGTGAGGATGAAAAGATTTATGAA	23571
Qy	901	ATTGTCTGCGCTCCATGTGTCATCATCATGAAATATCTAGACTGGATTAAGACGATTAACGAT	960
Db	23572	ATTGTCTGCGCTCCAAATGTGTCATCATCATGAAATATCTAGACTGGATTAAGACGATTAACGAT	23631
Qy	961	AAG 963	
Db	23632	AAG 23634	

RESULT 7	
LOCUS	ATH528171 650 bp DNA linear PLN 29-MAR-2003
DEFINITION	Arabidopsis thaliana T-DNA flanking sequence, left border, clone 157C08.
ACCESSION	AJ528171
VERSION	AJ528171.1 GI:26796431
KEYWORDS	left border; T-DNA flanking sequence.
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1
AUTHORS	Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F., Chauvin,S., Bechtold,N., Craud,C., Dehose,R., Pelletier,G., Lepoint,L., Caboche,M. and Lecharny,A.
TITLE	T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
JOURNAL	EMBO Rep. 3 (12), 1152-1157 (2002)
MEDLINE	22363535
PUBMED	12446565
REFERENCE	2 (bases 1 to 650)
AUTHORS	Balzergue,S.
TITLE	Direct Submission
JOURNAL	Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT	PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <a href="http://dbsgap.versailles.inra.fr/publiclines/">http://dbsgap.versailles.inra.fr/publiclines/</a> . This sequence has been generated in the framework of the French plant genomics program 'Genoplante' ( <a href="http://www.genoplante.com">http://www.genoplante.com</a> and <a href="http://genoplante-info.inbiohogen.fr">http://genoplante-info.inbiohogen.fr</a> ).
FEATURES	Location/Qualifiers
source	1..650
	/organism="Arabidopsis thaliana"
	/mol_type="genomic DNA"
	/cultivar="Massillaews-k1a"
	/db_xref="taxon:3702"
	/clone="157C08"
	/clone_1lb="Arabidopsis thaliana T-DNA insertion lines"
misc_feature	1..650
	/note="T-DNA flanking sequence
	left border"
ORIGIN	
Query Match	12.7%; Score 239; DB 8; Length 650;
Match Local Similarity	99.7%; Pred. No. 9.9e-111;
Matches 289; Conservative	0; Mismatches 1; Indels 0; Gaps 0;

OY		962	AAGGCGCTGGCATGGCAACGAAGACTGTGTGAAATATTTCGACAAGTATGAGGCCTTA	1021	
Db		303	AGTGCGCTGGCATTGGGCAACGAAGACTGTGTGAAATATTTCGACAAGTATGAGGCCTTA	362	
OY		1022	GAGCAGGCCATTCTTANAGTCACAGAGGCATGTGGAGTAGTGTTCTGATGTTTGAGA	1081	
Db		363	GAGCAGGCCATTCTTANAGTCACAGAGGCATGTGGAGTAGTGTTCTGATGTTTGAGA	422	
OY		1082	GCAGTGGCCACTGTGCTATTGTTGAGGCCGACAGCCTCCACGGGAGTTAGCTGAATGGGGT	1141	
Db		423	GCAGTGGCCACTGTGCTATTGTTGAGGCCGACAGCCTCCACGGGAGTTAGCTGAATGGGGT	482	
OY		1142	TAGATTAGAAATGTCCTGGGGGTGAGAAGCGCAGTATGTTTTCTGAGAGTGTGGCCAATGT	1201	
Db		483	TAGATTAGAAATGTCCTGGGGGTGAGAAGCGCAGTATGTTTTCTGAGAGTGTGGCCAATGT	542	
OY		1202	ATGGCTTCTTTCGACGAGAGCAATCTGACATTTTCATTCATCAACACTCT	1251	
Db		543	ATGGCTTCTTTCGACGAGAGCAATCTGACATTTTCATTCATCAACACTCT	592	
RESULT 8					
AC121725/c		178928 bp	DNA	linear HTG 11-OCT-2002	
LOCUS		Rattus norvegicus clone CHZ30-32812,	*** SEQUENCING IN PROGRESS		
DEFINITION		*** 3 unoverlaped pieces.			
ACCESSION	AC121725	GI:23804926			
VERSION	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.				
KEYWORDS	Rattus norvegicus (Norway rat)				
SOURCE	Rattus norvegicus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
REFERENCE	1 (bases 1 to 178928)				
AUTHORS	Muzny,D.,Marie,E.,Metzker,M.Lee,A.,Abramzon,S.,Adams,C.,Alder,J.,Allen,C.,Allen,H.,Alsbrooks,S.,Amin,A.,Anguelano,D.,Anyalobehnt,V.,Aoyagi,A.,Ayodeji,M.,Baca,E.,Baden,H.,Baldwin,D.,Bandaransaike,D.,Barber,M.,Barnstead,M.,Benahmed,F.,Blaisio,K.,Blair,J.,Blankenburg,K.,Blych,P.,Brown,M.,Bryant,N.,Bukey,C.,Burich,P.,Buttelli,K.,Calderon,E.,Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,H.,Center,A.,Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,Clelland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Greer,A.,D'Souza,L.,Devila,M.L.,Davis,C.,Day-Carroll,L.,De Anda,C.,Deedrich,D.,Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dimh.H.,Diya,K.,Draper,H.,Dugan-Rocha,S.,Dunn,A.,Durbin,K.,Duval,B.,Eaves,K.,Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,Fernandez,S.,Finley,M.,Flaggs,N.,Forbes,L.,Foster,M.,Poster,P.,Fernandez,S.,Gabisi,A.,Ganta,R.,Garcia,A.,Garner,T.,Garza,M.,Fraser,C.M.,Gabisi,A.,Ganta,R.,Grady,M.,Guerra,M.,Guevara,W.,Georgescu,P.E.,Geier,K.,Gill,R.,Grady,M.,Guerra,M.,Guevara,W.,Gunaratne,P.,Haaland,W.,Haml,C.,Hamilton,C.,Hamilton,K.,Harvey,J.,Havlak,P.,Hawes,A.,Henderson,N.,Hernandez,J.,Hernandez,R.,Hines,S.,Hiadun,S.L.,Hodgson,A.,Hogues,M.,Hollins,B.,Howells,S.,Huilyk,S.,Hume,J.,Idlebird,D.,Jackson,A.,Jackson,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Joilvet,A.,Kapathy,S.,Kelly,S.,Kelly,S.,Khan,Z.,King,L.,Kovar,C.,Kovats,C.,Kraft,C.L.,Lebow,H.,Lewan,U.,Lewis,L.,Li,Z.,Liu,J.,Liu,J.,Liu,W.,Liu,Y.,London,P.,Longacre,S., Lopez,J.,Lorenshewa,L.,Louisseg,H.,Lozada,R.J.,Lu,X.,Ma,J.,Maheshwari,M.,Mahnidartene,M.,Mahmoud,M.,Malloy,K.,Mangum,A.,Manushbari,M.,Martin,K.,Martín,R.,Martínez,E.,Mangum,B.,Mapua,P.,Martin,K.,Martín,R.,Martínez,E.,Mangum,B.,McLeod,M.P.,McNeill,T.Z.,Meenen,E.,Manning,S.,McLeod,M.P.,McNeill,T.Z.,Montemayor,J.,Moore,S.,Milosavljevic,A.,Miner,G.,Munja,E.,Montemayor,J.,Morjan,M.,Morris,K.,Morris,S.,Munidaas,M.,Murphy,M.,Nair,L.,Nankervis,C.,Neal,D.,Newton,G.,Olamunsgoon,A.,Pal,S.,Parke,K.,Nwaekelemehe,O.,Okwuonu,G.,Olarnunsgoon,A.,Pal,S.,Parke,K.,Plesnerak,S.,Paul,H.,Peres,A.,Peres,L.,Pfannkuch,C.,Plapper,F.,Poindexter,A.,Popovic,D.,Primus,E.,Pu,J.-L.,Purao,M.,Quiroz,J.,Rachelin,E.,Reeves,K.,Regler,M.A.,Reigh,R.,Ren,Y.,Reuter,M.,Reuter,M.,Richards,S.,Riggs,P.,River,C.,Rodney,T.,Rojs,A.,Rose,M.,Rose,R.,Ruiz,S.U.,				

SANDERS, W., SAVERY, G., SCHNER, S., SCOTT, G., SHATMAN, S., SHEN, H.,  
 SHETTY, J., SHVARTSBEGUN, A., SIBSON, I., SITTER, C.D., SNAJS, D.,  
 SNEED, A., SODERGREN, E., SONG, X.-Z., SORELLE, R., SOSSA, J.,  
 STEJLER, M., THOMAS, R., TUTTON, A., SVATEK, A., TABO, P., TAYLOR, C.,  
 TAYLOR, T., THOMAS, N., THOMAS, S., TINGEY, A., TREJO, Z., USMANT, K.,  
 VALS, R., VERA, V., VILLASANA, D., WALDRON, L., WALKER, B., WANG, J.,  
 WANG, Q., WANG, S., WARREN, J., WARREN, R., WLECZYK, R., WOODEN, H., WORLEY, K.,  
 WILLIAMS, G., WILSON, R., WLECZYK, R., WOODEN, H., WORLEY, K.,  
 WRIGHT, D., WRIGHT, R., WU, J., TAKUB, S., YEN, J., YOON, L., YOON, V.,  
 YU, F., ZHANG, J., ZHOU, J., ZHOU, X., ZHAO, S., DUNN, D., VON  
 NIEDERHUSEN, A., WEISS, R., SMITH, D.R., HOLT, R.A., SMITH, H.O.,  
 WEINSTECK, G. and Gibbs, R.A.  
 -----  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 -----  
 Direct Submission  
 Submitted (11-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 178928)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (11-OCT-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Oct 11, 2002 this sequence version replaced gi:21909375.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 -----  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GYUD  
 Center clone name: CH230-32812  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 143032 bases at least Q40  
 Consensus quality: 146972 bases at least Q30  
 Consensus quality: 149332 bases at least Q20  
 Estimated insert size: 145820; sum-of-contigs estimation  
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 175436: contig of 175436 bp in length  
 \* 175437 175536: gap of unknown length  
 \* 175537 176579: contig of 1043 bp in length  
 \* 176580 176679: gap of unknown length  
 \* 176680 178928: contig of 2249 bp in length.  
 Location/Qualifiers  
 1. 178928  
 /organism="Rattus norvegicus"  
 FEATURES  
 source

```

misc_feature      /mol_type="genomic DNA"
                  /db_xref="taxon:10116"
                  /clone="CH230-32812"
                  1..1169
                  /note="wgs_end_extension
clone_end:Sp6"
misc_feature      27453..29430
                  /note="wgs_end_extension
clone_end:Sp6"
misc_feature      77608..79432
                  /note="wgs_end_extension
clone_end:Sp6"
misc_feature      99609..100719
                  /note="wgs_end_extension
clone_end:Sp6"
misc_feature      complement(118199..119034)
                  /note="clone_boundary
clone_end:Sp6
misc_feature      end:sequence:RXAOM49TV"
                  125929..127858
                  /note="wgs_contig"
misc_feature      156410..157915
                  /note="wgs_contig"
misc_feature      complement(172023..172895)
                  /note="clone_boundary
clone_end:T7
misc_feature      site:Mbol
                  end:sequence:RXAOM49TV"
                  173824..175436
                  /note="wgs_end_extension
clone_end:T7"

```

## ORIGIN

```

Query Match      1.2%; Score 22; DB 2; Length 178928;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 492 TTCTGAGAGAGATGATTC 513
Db 112510 TTCTGAGAGAGATGATTC 112489

```

```

RESULT 9
AL928719/c      197909 bp      DNA      linear      ROD 28-JAN-2003
LOCUS           Mouse DNA sequence from clone RP23-419G21 on chromosome 2, complete
DEFINITION
ACCESSION      AL928719
VERSION        AL928719.6
KEYWORDS       HTG.
SOURCE         Mus musculus (house mouse)
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
                1 (bases 1 to 197909)
REFERENCE      1 (bases 1 to 197909)
AUTHORS       Leongamornlert, D.
TITLE         Direct Submision
JOURNAL        Submitted (28-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
                Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                humquerry@sanger.ac.uk
                On Nov 3, 2002 this sequence version replaced gi:24395353.
                Sequence from the Mouse Genome Sequencing Consortium whole genome
                shotgun may have been used to confirm this sequence. Sequence data
                from the whole genome shotgun alone has only been used where it has
                a phred quality of at least 30.
                ----- Genome Center
                Center: Wellcome Trust Sanger Institute
                Center code: SC
                Web site: http://www.sanger.ac.uk
                Contact: humquerry@sanger.ac.uk
                -----
                During sequence assembly data is compared from overlapping clones.

```

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL, Sw., SWISSPROT, Tr., TrEMBL, Wp., WormPep; Information on the WormPep database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-419G21 is from the RPI-23 Mouse BAC Library.

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.

## FEATURES

```

source
Location/Qualifiers
1..197909
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-419G21"
/clone_11b="RPI-23"

```

## ORIGIN

```

Query Match      1.2%; Score 22; DB 10; Length 197909;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 567 TGTGACTCAAGAGCCATGCA 588
Db 34650 TGTGACTCAAGAGCCATGCA 34629

```

```

RESULT 10
AC078933      214166 bp      DNA      linear      HTG 11-AUG-2000
LOCUS           Mus musculus chromosome 5 clone RP23-201E13 strain C57BL6/J,
DEFINITION      WORKING DRAFT SEQUENCE, 17 unordered pieces.
ACCESSION      AC078933
VERSION        AC078933.1
KEYWORDS       HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE         Mus musculus (house mouse)
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
                1 (bases 1 to 214166)
REFERENCE      1 (bases 1 to 214166)
AUTHORS       Bouffard, G.G., Dietrich, N.L., Bagle, W.O., Gupta, J., Ho, S.-L.,
                Huang, M.C., Idol, J., Lee-Lin, S.-O., Maduro, O.L., Maduro, V.B.,
                Maestrian, S.D., McCloskey, J.C., McDowell, J., Ojodu, M.A., Pearson, R.,
                Stanciripop, S., Summers, T.J., Thomas, J.W., Thomas, P.J.,
                Tlionson, E.E., Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A.,
                Welteby, K.D., and Green, E.D.
TITLE         NISC Mouse Sequencing Initiative
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 214166)
AUTHORS       Green, E.D.
TITLE         Direct Submision
JOURNAL        Submitted (11-AUG-2000) NIH Intramural Sequencing Center, 8717
                Grovemont Circle, Gaithersburg, MD 20877, USA
                ----- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_mouse@nhgri.nih.gov
                -----

```

```

----- Project Information
Center project name: vd
Center clone name: 201E13
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 204605 bases at least Q40
Consensus quality: 207891 bases at least Q30
Consensus quality: 209797 bases at least Q20
Insert size: 218000; agarose-gel
Insert size: 237000; pulse-field-gel
Insert size: 212586; sum-of-contigs
Quality coverage: 4.82x in Q20 bases; agarose-gel
Quality coverage: 4.43x in Q20 bases; pulse-field-gel
Quality coverage: 4.94x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 3207: contig of 3207 bp in length
3208 3307: gap of unknown length
3308 7622: contig of 4315 bp in length
7623 7722: gap of unknown length
7723 13087: contig of 5365 bp in length
13088 13187: gap of unknown length
13188 20448: contig of 7261 bp in length
20449 20549: gap of unknown length
20549 26559: contig of 6111 bp in length
26560 26760: gap of unknown length
26760 35349: contig of 8590 bp in length
35350 35449: gap of unknown length
35450 43812: contig of 8363 bp in length
43813 52486: contig of 8574 bp in length
43913 52586: gap of unknown length
52487 62100: contig of 9514 bp in length
52587 62200: gap of unknown length
62101 77345: contig of 15145 bp in length
77346 77445: gap of unknown length
77446 87428: contig of 9983 bp in length
87429 87528: gap of unknown length
87529 100606: contig of 13078 bp in length
100607 100707: gap of unknown length
100707 117904: contig of 17198 bp in length
117905 118004: gap of unknown length
118005 135715: contig of 17711 bp in length
135716 135815: gap of unknown length
135816 153408: contig of 17593 bp in length
153409 153508: gap of unknown length
153509 181134: contig of 27526 bp in length
181135 214186: gap of unknown length
181135 214186: contig of 33052 bp in length.
-----
FEATURES
source
1. 214186
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-201E13"
/clone_lib="RPCI mouse BAC library 23"
1. 3207
/misc_feature
/note="assembly_fragment"
3308. 7622
/misc_feature
/note="assembly_fragment"
7723. 13087
/misc_feature
/note="assembly_fragment"

```

```

misc_feature 13188..20448
/note="assembly_fragment"
misc_feature 20549..26559
/note="assembly_fragment"
misc_feature 26760..35349
/note="assembly_fragment"
misc_feature 35450..43812
/note="assembly_fragment"
misc_feature 43913..52486
/note="assembly_fragment"
misc_feature 52587..62100
/note="assembly_fragment"
misc_feature 62201..77345
/note="assembly_fragment"
misc_feature 77446..87428
/note="assembly_fragment"
misc_feature 87529..100606
/note="assembly_fragment"
misc_feature 100707..117904
/note="assembly_fragment"
misc_feature 118005..135715
/note="assembly_fragment"
misc_feature 135816..153408
/note="assembly_fragment"
misc_feature 153509..181034
/note="assembly_fragment"
misc_feature 181135..214186
/note="assembly_fragment"
-----
ORIGIN
Query Match 1.2%; Score 22; DB 2; Length 214186;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 567 TGTGAGTCAAAAGAGCCATGCA 588
Db 99680 TGTGAGTCAAAAGAGCCATGCA 99701
-----
RESULT 11
AC079182/c 218774 bp DNA linear HTG 23-AUG-2000
LOCUS Mus musculus chromosome 5 clone RP23-203F6 strain C57BL6/J, WORKING
DEFINITION DRAFT SEQUENCE, 16 unordered pieces.
ACCESSION AC079182
VERSION AC079182.1 GI:9886001
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 218774)
Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Bouffard, G.G., Dietrich, N.L., Bagle, W.O., Gupta, J., Ho, S.-L.,
Idol, J., Lee-Jin, S.-O., Legaspi, R., Lim, M., Maduro, Q.L.,
Maduro, V.B., Mastrian, S.D., McCloskey, J.C., McDowell, J.,
Pearson, R., Stantitrop, S., Sumner, T.J., Thomas, J.W., Thomas, P.J.,
Tongson, E.E., Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A.,
Wehrhery, K.D. and Green, E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 218774)
Green, E.D.
Direct Submission
Submitted (23-AUG-2000) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Galtersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
COMMENT
Center code: NISC

```

```

Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@hgr1.nih.gov
----- Project Information -----
Center project name: ve
Center clone name: 203P06
----- Summary Statistics -----
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Assembly quality: 208040 bases at least Q40
Consensus quality: 211445 bases at least Q30
Consensus quality: 213463 bases at least Q20
Insert size: 227000; agarose-gel
Insert size: 242000; pulse-field-gel
Insert size: 217274; sum-of-ctctigs
Quality coverage: 4.70x in Q20 bases; agarose-gel
Quality coverage: 4.40x in Q20 bases; pulse-field-gel
Quality coverage: 4.91x in Q20 bases; sum-of-ctctigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2001: contig of 2001 bp in length
2 2002 2101: gap of unknown length
3 2102 4870: contig of 2769 bp in length
4 4871 4970: gap of unknown length
5 4971 8469: contig of 3499 bp in length
6 8470 8569: gap of unknown length
7 8570 14232: contig of 5663 bp in length
8 14233 14332: gap of unknown length
9 14333 21631: contig of 7299 bp in length
10 21632 21731: gap of unknown length
11 21732 29661: contig of 7930 bp in length
12 29662 29761: gap of unknown length
13 29762 38142: contig of 8381 bp in length
14 38143 38242: gap of unknown length
15 38243 47730: contig of 9488 bp in length
16 47731 47831: gap of unknown length
17 47831 55944: contig of 8114 bp in length
18 55945 56044: gap of unknown length
19 56045 70548: contig of 14504 bp in length
20 70549 70648: gap of unknown length
21 70649 85515: contig of 14867 bp in length
22 85516 85615: gap of unknown length
23 85616 103358: contig of 17743 bp in length
24 103359 103458: gap of unknown length
25 103459 125084: contig of 21626 bp in length
26 125085 125184: gap of unknown length
27 125185 153356: contig of 28172 bp in length
28 153357 153456: gap of unknown length
29 153457 184845: contig of 31389 bp in length
30 184846 184945: gap of unknown length
31 184946 218774: contig of 33829 bp in length.
Location/Qualifiers
1. 218774
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone_lib="RP23-203P6"
/clone_lib="PPCI mouse BAC library 23"
1. 2001
/note="assembly_fragment"
2102. 4870
/note="assembly_fragment"
4971. 8469
/note="assembly_fragment"

```

```

misc_feature 8570..14232
/note="assembly_fragment"
misc_feature 1433..21631
/note="assembly_fragment"
misc_feature 21732..29661
/note="assembly_fragment"
misc_feature 29762..38142
/note="assembly_fragment"
misc_feature 38243..47730
/note="assembly_fragment"
misc_feature 47831..55944
/note="assembly_fragment"
misc_feature 56045..70548
/note="assembly_fragment"
misc_feature 70649..85515
/note="assembly_fragment"
misc_feature 85616..103358
/note="assembly_fragment"
misc_feature 103459..125084
/note="assembly_fragment"
misc_feature 125185..153356
/note="assembly_fragment"
misc_feature 153457..184845
/note="assembly_fragment"
misc_feature 184946..218774
/note="assembly_fragment"
ORIGIN
Query Match 1.2%; Score 22; DB 2; Length 218774;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 567 TGTGACTCAAAAGAGCCATGCA 588
Db 76509 TGTGACTCAAAAGAGCCATGCA 76488
RESULT 12
AC115306/c 236508 bp DNA linear HTG 23-NOV-2002
LOCUS Rattus norvegicus clone CH230-11F1, WORKING DRAFT SEQUENCE, 3
DEFINITION
unordered pieces.
ACCESSION
AC115306.4 GI:25188797
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 236508)
REFERENCE
1. Wuzny,D,Marie., Metzker,M,lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguitano,D.,
Anyalebech,V., Ayvagi,A., Ayodeji,M., Baca,B., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Bunay,C., Burch,P., Butrell,K., Calderon,E.,
Bryant,N., Buhay,J., Blankenburg,K., Blyth,P., Brown,M.,
Cardenas,V., Carter,K., Cavazos,I., Casas,R., Chatterjee,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Derramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebreyorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunatune,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,U.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulik,S., Hume,J., Idlebird,D., Jackson,A.,

```

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, C., Kows, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshew, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martineau, E., Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munda, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, G., Olarnungsoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plummer, F., Polindexter, A., Popovic, D., Primm, E., Pu, L., Puzos, M., Quirio, J., Rachlin, E., Reeves, R., Regier, M.A., Reigh, R., Riley, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Shen, H., Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Sney, J., Snavarsky, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J., Steimle, M., Strong, R., Sutton, A., Swatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trevis, Z., Usmani, K., Valas, R., Vera, V., Villaseca, D., Waldron, J., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Woodson, H., Wortley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, X., Dunn, D., von Weinstock, G., and Gibbs, R.A.

# TITLE JOURNAL AUTHORS TITLE JOURNAL

2 (bases 1 to 236508)  
Unpublished  
Direct Submission  
Submitted (17-Mar-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 236508)  
Rat Genome Sequencing Consortium.  
Submitted (23-Nov-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## REFERENCE AUTHORS TITLE JOURNAL

On Nov 23, 2002 this sequence version replaced g1:22772519.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Alas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a contig-scaffold). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GBAS  
Center clone name: CH230-11F1  
----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 219730 bases at least Q40  
Consensus quality: 222791 bases at least Q30  
Consensus quality: 224323 bases at least Q20  
Estimated insert size: 225524; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\*\*\*\*\* NOTE: Estimated insert size may differ from sequence length

## FEATURES source

(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
1 125318: contig of 125318 bp in length  
\* 125319 125418: gap of unknown length  
\* 125419 180275: contig of 54857 bp in length  
\* 180276 180375: gap of unknown length  
\* 180376 236508: contig of 56133 bp in length.  
Location/Qualifiers  
1. 236508 "Rattus norvegicus"  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-11F1"  
1. 1641  
/note="wgs contig"  
155976..157691  
/note="wgs contig"  
233371..235181  
/note="wgs contig"  
235232..236508  
/note="wgs contig"

## ORIGIN

Query Match 1.2%; Score 22; DB 2; Length 236508;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 AACGAGAAACCTTCGAAA 182  
Db 162467 AACGAGAAACCTTCGAAA 162446

RESULT 13  
AC140364 280016 bp DNA linear HTG 05-NOV-2003  
LOCUS  
DEFINITION  
SEQUENCE, 8 unordered pieces.  
ACCESSION  
AC140364.2 GI:38176012  
VERSION  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS  
Mus musculus (house mouse)  
SOURCE  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (05-NOV-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
On Nov 5, 2003 this sequence version replaced g1:28475622.

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu  
Contact: submissions@wustl.edu  
----- Project Information



Center project name: M.BB0075K05

## ----- Summary Statistics -----

Sequencing vector: M13; 0%  
 Chemistry: Dye-Primer ET; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 278028 bases at least Q40  
 Consensus quality: 278887 bases at least Q30  
 Consensus quality: 279492 bases at least Q20

NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1202: contig of 1202 bp in length  
 \* 1203 1302: gap of unknown length  
 \* 1303 3866: contig of 2564 bp in length  
 \* 3867 3966: gap of unknown length  
 \* 3967 24768: contig of 20802 bp in length  
 \* 24769 24868: gap of unknown length  
 \* 24869 48433: contig of 23565 bp in length  
 \* 48434 48533: gap of unknown length  
 \* 48534 85803: contig of 37270 bp in length  
 \* 85804 85903: gap of unknown length  
 \* 85904 127685: contig of 41782 bp in length  
 \* 127686 127785: gap of unknown length  
 \* 127786 198220: contig of 70435 bp in length  
 \* 198221 198320: gap of unknown length  
 \* 198321 280016: contig of 81696 bp in length.  
 Location/Qualifiers

## FEATURES

source  
 1. .280016  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="UNK"  
 /clone="RP24-75K5"  
 1. .1202  
 /note="assembly\_name:Contig20"  
 misc\_feature 1303..3866  
 /note="assembly\_name:Contig22"  
 misc\_feature 3867..24768  
 /note="assembly\_name:Contig23"  
 misc\_feature 24869..48433  
 /note="assembly\_name:Contig24"  
 misc\_feature 48534..85803  
 /note="assembly\_name:Contig25"  
 misc\_feature 85904..127685  
 /note="assembly\_name:Contig26"  
 misc\_feature 127786..198220  
 /note="assembly\_name:Contig27"  
 misc\_feature 198321..280016  
 /note="assembly\_name:Contig28"

## ORIGIN

Query Match 1.2%; Score 22; DB 2; Length 280016;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 567 TGTGAGTCAAAAGAGCCATGGA 588  
 |||||  
 Db 58633 TGTGAGTCAAAAGAGCCATGGA 58654

RESULT 14  
 AC073813 304407 bp DNA linear HTG 29-JUN-2000  
 LOCUS AC073813  
 DEFINITION Mus musculus clone RP23-69B1, WORKING DRAFT SEQUENCE, 58 unordered

ACCESSION AC073813  
 VERSION AC073813.1 GI:8910430  
 KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE Bukaryoca; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 304407)  
 TITLE DOE Joint Genome Institute.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 304407)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

## COMMENT

Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov  
 Project Information  
 Center Project Name: 1763393  
 Center clone name: RPCI-23\_69B1

-----  
 Summary Statistics  
 Consensus quality: 257140 bases at least Q40  
 Consensus quality: 282847 bases at least Q30  
 Consensus quality: 288639 bases at least Q20  
 Estimated insert size: 272860; agarose-fp estimation  
 Estimated insert size: 298707; sum-of-contigs estimation  
 Quality coverage: 9.46 in Q20 bases; agarose-fp estimation  
 Quality coverage: 8.64 in Q20 bases; sum-of-contigs estimation.  
 NOTE: This is a 'working draft' sequence. It currently consists of 58 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
 1 1167: contig of 1167 bp in length  
 \* 1168 1267: gap of unknown length  
 \* 1268 2281: contig of 1014 bp in length  
 \* 2282 2381: gap of unknown length  
 \* 2382 3767: contig of 1366 bp in length  
 \* 3768 3867: gap of unknown length  
 \* 3868 5372: contig of 1505 bp in length  
 \* 5373 5472: gap of unknown length  
 \* 5473 6484: gap of unknown length  
 \* 6485 6594: gap of unknown length  
 \* 6595 7909: contig of 1315 bp in length  
 \* 7910 8009: gap of unknown length  
 \* 8010 9416: contig of 1407 bp in length  
 \* 9417 9516: gap of unknown length  
 \* 9517 10776: contig of 1260 bp in length  
 \* 10777 10876: gap of unknown length  
 \* 10877 12279: contig of 1403 bp in length  
 \* 12280 12379: gap of unknown length  
 \* 12380 13920: contig of 1541 bp in length  
 \* 13921 14020: gap of unknown length  
 \* 14021 14917: contig of 1337 bp in length  
 \* 14918 15417: gap of unknown length  
 \* 15418 15517: gap of unknown length  
 \* 15518 16897: contig of 1380 bp in length  
 \* 16898 16997: gap of unknown length  
 \* 16998 18940: contig of 1943 bp in length  
 \* 18941 18941: gap of unknown length  
 \* 18942 20880: contig of 1850 bp in length  
 \* 20881 20990: gap of unknown length  
 \* 20991 22965: contig of 1975 bp in length  
 \* 22966 23065: gap of unknown length  
 \* 23066 24501: contig of 1436 bp in length



24502 24601: gap of unknown length  
 24602 26428: contig of 1827 bp in length  
 26429 26528: gap of unknown length  
 26529 28503: contig of 1975 bp in length  
 28504 28603: gap of unknown length  
 28604 30124: contig of 1521 bp in length  
 30125 30224: gap of unknown length  
 30225 32292: contig of 2068 bp in length  
 32293 32393: gap of unknown length  
 32394 34644: contig of 2252 bp in length  
 34645 34744: gap of unknown length  
 34745 37728: contig of 2984 bp in length  
 37729 37828: gap of unknown length  
 37829 40230: contig of 2402 bp in length  
 40231 40330: gap of unknown length  
 40331 42294: contig of 1964 bp in length  
 42295 42394: gap of unknown length  
 42395 44875: contig of 2481 bp in length  
 44876 44975: gap of unknown length  
 44976 48629: contig of 3654 bp in length  
 48630 48729: gap of unknown length  
 48730 51512: contig of 2783 bp in length  
 51513 51612: gap of unknown length  
 51613 54010: contig of 2398 bp in length  
 54011 54110: gap of unknown length  
 54111 55956: contig of 1846 bp in length  
 55957 55956: gap of unknown length  
 55957 57509: contig of 1453 bp in length  
 57510 57609: gap of unknown length  
 57610 63737: contig of 6128 bp in length  
 63738 63837: gap of unknown length  
 63838 66881: contig of 3044 bp in length  
 66882 66981: gap of unknown length  
 66982 70918: contig of 3937 bp in length  
 70919 71018: gap of unknown length  
 71019 75925: contig of 4907 bp in length  
 75926 76025: gap of unknown length  
 76026 78940: contig of 2915 bp in length  
 78941 79040: gap of unknown length  
 79041 84797: contig of 5757 bp in length  
 84798 84898: gap of unknown length  
 84899 88965: contig of 4068 bp in length  
 88966 89065: gap of unknown length  
 89066 92420: contig of 3355 bp in length  
 92421 92520: gap of unknown length  
 92521 97335: contig of 4815 bp in length  
 97336 97435: gap of unknown length  
 97436 102886: contig of 5451 bp in length  
 102887 102986: gap of unknown length  
 102987 106824: contig of 3838 bp in length  
 106825 106924: gap of unknown length  
 106925 114199: contig of 7275 bp in length  
 114200 114299: gap of unknown length  
 114300 120390: contig of 6691 bp in length  
 120391 121090: gap of unknown length  
 121091 126531: contig of 8441 bp in length  
 126532 126631: gap of unknown length  
 126632 134893: contig of 5262 bp in length  
 134894 134993: gap of unknown length  
 134994 142772: contig of 7779 bp in length  
 142773 142872: gap of unknown length  
 142873 157121: contig of 14249 bp in length  
 157122 157221: gap of unknown length  
 157222 164701: contig of 7480 bp in length  
 164702 164801: gap of unknown length  
 164802 171481: contig of 6680 bp in length  
 171482 171581: gap of unknown length  
 171582 184008: contig of 12427 bp in length  
 184009 184108: gap of unknown length  
 184109 192382: contig of 8284 bp in length  
 192393 192492: gap of unknown length  
 192493 203020: contig of 10528 bp in length  
 203021 203120: gap of unknown length

203121 211871: contig of 8751 bp in length  
 211872 223920: contig of 11949 bp in length  
 223921 224020: gap of unknown length  
 224021 237820: contig of 13800 bp in length  
 237821 237920: gap of unknown length  
 237921 256627: contig of 18707 bp in length  
 256628 256727: gap of unknown length  
 256728 280655: contig of 23928 bp in length  
 280656 280755: gap of unknown length  
 280756 304407: contig of 23652 bp in length.  
 Location/Qualifiers  
 1. 304407  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP23-69B1"  
 /clone\_id="RP23 mouse BAC library 23"

ORIGIN  
 Query Match 1.2% Score 22; DB 2; Length 304407;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 GCTCAGCACAGCTGTCAGG 457  
 |||||  
 Db 5763 GCTCAGCACAGCTGTCAGG 5784

RESULT 15  
 AE005048  
 LOCUS  
 Halobacterium sp. NRC-1 section 79 of 170 of the complete genome.  
 DEFINITION  
 AE005048 AE004437  
 ACCESSION  
 AE005048.1 GI:10580743  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Halobacterium sp. NRC-1  
 Halobacterium sp. NRC-1  
 Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;  
 Halobacteriaceae; Halobacterium.  
 1 (bases 1 to 10225)  
 Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,  
 Shukla,H.D., Lasky,S.R., Baliga,N., Thorson,V., Shrogha,J.,  
 Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,  
 Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,  
 Maddock,D.G., Jablonski,P.E., Krebs,M.P., Angeline,C.M., Dale,H.,  
 Isebnarger,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,  
 Jung,K.H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,  
 Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and  
 Dassa,S.  
 Genome sequence of Halobacterium species NRC-1  
 Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)  
 20504483  
 MEDLINE  
 PUBMED  
 11016950

REFERENCE  
 AUTHORS  
 2 (bases 1 to 10225)  
 Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,  
 Shukla,H.D., Lasky,S.R., Baliga,N., Thorson,V., Shrogha,J.,  
 Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,  
 Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,  
 Maddock,D.G., Jablonski,P.E., Krebs,M.P., Angeline,C.M., Dale,H.,  
 Isebnarger,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,  
 Jung,K.H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,  
 Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and  
 Dassa,S.  
 Direct Submission  
 Submitted (14-JUL-2000) Institute for Systems Biology, 4225  
 Roosevelt Way NE, Seattle, WA 98105, USA  
 Location/Qualifiers  
 1. 10225  
 /organism="Halobacterium sp. NRC-1"  
 /mol\_type="genomic DNA"  
 /strain="NRC-1"  
 /db\_xref="taxon:64091"

FEATURES  
 source

gene 125. .1510  
/gene="huth"  
/note="synonym: VNG1212G"  
CDS 125. .1510  
/gene="huth"  
/note="huth"  
/codon\_start=1  
/transl\_table=11  
/product="histidine ammonia-lyase"  
/protein\_id="AAG19580.1"  
/db\_xref="GI:10580744"  
/translation="MSDTRIDAAADREALQANLVRSAGAGSELDTPAAVRALLVTREL  
ALAKYSGIRERVLDTVLGLNIGVPPVPSRGLGASGDLPIAMSRVLGEGQAD  
VGERMPAAELAAADLEPTLQAKGALINGTOLTGYAALALVDAEVRASADTA  
GALTTEVMTTASGAPATHEVRPHGQASAHIRNLITGASEVLHHDCRVDQAY  
SIRCLPQVIGAVARDALDHLRAVATELNSDTPVYFAGDRVSPRSGCDRAVVS  
NFEGLVALRLGTAASALAEALAISSRTDRLLNPEPTEPHLEPFLAPDSGLHSLMT  
PQYTAASLVNDLSLQPTLDNVAVSAGQEDHVSAGAAVNFREAVKAAVVGVEL  
LCGAQGREFLDPDLALGAGTAAAYDLVREVSFVAGDRALADMAAVGDLVRAGLVED  
AVARALDAPLA"  
1598. .2578  
/gene="VNG1213C"  
1598. .2578  
/gene="VNG1213C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1213C"  
/protein\_id="AAG19581.1"  
/db\_xref="GI:10580745"  
/translation="MSAAQOTVYHRSACRSGRASDIGNPLPTATFDRLLCHCATARR  
PTSECAPDTERASAPTRPEETALPMQOPQAVDMRWVNSFLPAAGVGETERLMOGV  
THMTDGDGPGIGDATTENIYGFIDAARALDAGNTSPFADLPSGLIMRLYENAD  
DVAFPLEITTTGDDQSSVYTVVSVHKGQDTTIVRGSUDLAEQLQLLADPMVYANG  
KRPDAFLHEHFDLSLDSPHLDLMPCKRDLDTGLKQVERDLGVRAEDVDGREA  
VRLMRRYASDNDDEAALDRLVTYNQDTQNLQALLDAVBERLHARDVEPHV"  
3236. .4366  
/gene="VNG1214H"  
3236. .4366  
/gene="VNG1214H"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1214H"  
/protein\_id="AAG19582.1"  
/db\_xref="GI:10580746"  
/translation="MTSHDYRLAEIEDGPLERLAIDLTLTRDYGCVDPQGGKQDG  
GKQGLLDGPDGNNIIYHVSREDWKQKLTDLGKAADHDVDYDIFVYVYNR.IITNQ  
KPVDAQPFVDEYGEIDIMDEGRLESLDNNHODLREYRLIADDEDPTGKTAII  
DEBLRLRRRDNELPRVBDAPYAVLHVPHDVSQDTPRRHDLPTPLSGRYGYS  
YENTLDGRVTYVAGGDNEDPGYIYVDTGCMVFAVDAFMGNENMTIGQSFKKLIGD  
ATBYGNALNELELDGPFVGLSVLSVKYSFPTKKGCGGDRNGPVPFRNDIEAKPH  
TYEEDTSPTEGAARKGFPDRVWARGARWSDSPYSDNEWHEFR"  
complement(4565. .4638)  
/gene="trn19"  
complement(4565. .4638)  
/gene="trn19"  
/product="tRNA-Thr"  
complement(4650. .5132)  
/gene="pai1"  
/note="synonym: VNG1215G"  
complement(4650. .5132)  
/gene="pai1"  
/note="pai1"  
/codon\_start=1  
/transl\_table=11  
/product="sporulation regulator homolog"  
/protein\_id="AAG19583.1"  
/db\_xref="GI:10580747"  
/translation="MTVRPASGTDIDAVVAELMWQLAEQRAHSGSHLLAENRAQADL  
IGQYVADGVAVTQAGRTVGFWFHTETGFIETDRGTIDLVYAPGKGGKRSQA  
LITHARELRSRGADALAIETALDNNAAARLYERRGYAPHRVTFEOSVENDTHSKDDP  
"

gene complement(5129. .6325)  
/gene="pgk"  
/note="synonym: VNG1216G"  
CDS complement(5129. .6325)  
/gene="pgk"  
/note="pgk"  
/codon\_start=1  
/transl\_table=11  
/product="glucose-6-phosphate isomerase"  
/protein\_id="AAG19584.1"  
/db\_xref="GI:10580748"  
/translation="MAIRFLDLDLAAANRAIGVRVDINSPLTAAGGLADARLRPHVDT  
LAEILDAARVAVLVAHQRPQGDSEFARLERHADRLDALDAEVSYCDATFSGARAV  
ALDAPEAVLLETRFYSSEYVAFAPERAADTALVDGLAPALDAVYNDAPFAAHRQAP  
SLVGPEPVYASGRVMEALDLSGVDTPPTPRVYVGGAARVPSVEVAALASHGL  
ANNVLTGVAVAVPLAATGVDDIGRSTDEIHERDVGTEIARAADLLAANNDAHLHPVD  
VAVRGAACEISTDLPPAGAEAVCDIGSDTVADADADVLADETVVVGPRGVEFDD  
LEADRTGVCFDASEVESHIVGGGDTAAIRRFDTITGFHVSTGGGAILNLTDADLP  
AVAAALR"  
complement(6428. .7144)  
/gene="VNG1218C"  
complement(6428. .7144)  
/gene="VNG1218C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1218C"  
/protein\_id="AAG19585.1"  
/db\_xref="GI:10580749"  
/translation="MQRFAGCTPTAVMAVALGPPRPSVHGFTYRGLLYRRMNDVRLSE  
ALSASRVVSAARAAPDAATTLREHDATTAVQAGDLGVTADAVDAVADGHLSD  
STVADVMAVDTEVSLSDATVADADIDAMARTGDCVLVVDADAELAGVSAOAILALA  
LANNDRGSRGRLPGGDPDPVGAAGSAGAPDPPESRDGPTTAGANTYSYSGVCERGG  
LAAALGRVNRRLCTECSRSL"  
7107. .7853  
/gene="urk"  
/note="synonym: VNG1219G"  
7107. .7853  
/gene="urk"  
/note="urk"  
/codon\_start=1  
/transl\_table=11  
/product="uridine kinase"  
/protein\_id="AAG19586.1"  
/db\_xref="GI:10580750"  
/translation="MTAVGVQPAKRCIQHASPFPRAVSMITPFAIGIAGTGAQKTV  
AEETIDNVGESATLPLDNYEDLSDRPPEERANANYDPSAFEWELRTHIDALISG  
QSIEMEQVDPEHVRKADRVVPEPDVIVLEGIALLSDETVNDMLDHLHYVEDADVR  
ILRLREVDVFERHGRELEGVMDQYLSTVKMHQGFLEPTGRHADIIIPEGANSVAVML  
EKVQAESEMTAMARDDTAFEDATVYNQ"  
complement(7911. .8234)  
/gene="VNG1220H"  
complement(7911. .8234)  
/gene="VNG1220H"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1220H"  
/protein\_id="AAG19587.1"  
/db\_xref="GI:10580751"  
/translation="MAEMEDPSGRDWHDPDGEHSGEGKRYGLAVLSKWTVEGEDFP  
LDAEFVAEYGEYVPLNLNGTVSVADIFEHILDABEFETKVGFKHTVGAALDADPDL  
YHPDA"  
complement(8533. .8615)  
/gene="trn20"  
complement(8533. .8615)  
/gene="trn20"  
/product="tRNA-Ser"  
8926. .10119  
/gene="orc8"  
/note="synonym: VNG1224G"  
8926. .10119  
/gene="orc8"

```

/codon="orc8"
/codon_start=1
/translate="orc" / cell division control protein 6"
/product="orc" / protein_id="AAC19588.1"
/db_xref="GI:10580752"
/translation="MKPRFDRVVELPANKDIVLKHDEEELRERDEEIDYVNALQDV
VDGMEPRNFDVYGTGTGKGAIVRVMKALREYADRGVGSVTSVENGCHNPPSSQV
AAIALLVLEIRDDTSDLTLTGTSITVYMLALDELAEAGVLLYLDELIDLDDMLT
LYLPRAKATLINSIDSSVAVQCAALNADRSARADVLRESINVAIGERETVE
ILDPRAEALSSGLTSGVAVPCCALNADRSARADVLRESINVAIGERETVE
DVEEAIVARRVERGIKOSIKDLTTHGQVLLAVTQALADDPVPAKELYEYVADIAA

```

	Query Match	1.1%;	Score 21;	DB 1;	Length 10225;
	Best Local Similarity	100.0%;	Mismatches	0;	Gaps 0;
	Matches	21;	Conservative	0;	Indels 0;
QY	1104 GGGCGAAGCCTCCACCGGA	1124			
	2				
DB	2536 GGCGAAAGCCTTCACCGGA	2556			

Search completed: April 9, 2004, 05:51:30  
Job time : 4852.7 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2004, 16:25:39 ; Search time 479.978 seconds  
(without alignments)  
16621.841 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878  
Sequence: 1 atgagctcagggctgtcc.....ttcacatgaagatgatga 1878

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N.Geneseq\_29Jan04.\*

1: geneseqn1980s.\*  
2: geneseqn1990s.\*  
3: geneseqn2000s.\*  
4: geneseqn2001as.\*  
5: geneseqn2001bs.\*  
6: geneseqn2002s.\*  
7: geneseqn2003as.\*  
8: geneseqn2003bs.\*  
9: geneseqn2003cs.\*  
10: geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1878	100.0	1878	4	AAFP25374
2	963	51.3	3275	4	AAFP25373
3	24	1.3	27	4	AAFP25372
4	20	1.1	23	4	AAFP25371
5	20	1.1	596	4	AAFI2773
6	20	1.1	700	4	AAH93017
7	20	1.1	47475	3	AAAB81465
8	20	1.1	110000	3	AAAB81489_5
9	20	1.1	154902	6	ABQ88198
10	20	1.1	349980	3	AAFP21612
11	19	1.0	481	3	AAA68016
12	19	1.0	481	9	ADDA1766
13	19	1.0	484	4	AAK64246
14	19	1.0	485	2	AAV23847
15	19	1.0	495	2	AAZ06848
16	19	1.0	495	2	AAZ06848
17	19	1.0	495	3	AAA65997
18	19	1.0	495	3	AAA67934
19	19	1.0	495	3	ADDA1684
20	19	1.0	501	6	AAAG1626
21	19	1.0	538	6	ABT10660
22	19	1.0	1341	6	ABZ14660
23	19	1.0	1838	4	ABLI1339
24	19	1.0	1876	4	ABLI5067

24	19	1.0	2000	2	AAV33671	AAV33671 Candida k
25	19	1.0	2000	2	AAV33672	AAV33672 Candida k
26	19	1.0	2000	4	AAH73590	AAH73590 Candida k
27	19	1.0	2000	4	AAH73589	AAH73589 Candida k
28	19	1.0	2000	7	ACA62356	ACA62356 Gene enco
29	19	1.0	2000	7	ACA62357	ACA62357 mRNA enco
30	19	1.0	2000	8	ACA62409	ACA62409 C. krusei
31	19	1.0	2000	8	ACA62410	ACA62410 C. krusei
32	19	1.0	2195	4	AAK87051	AAK87051 Human imm
33	19	1.0	2204	4	ABL26576	ABL26576 Drosophila
34	19	1.0	2462	4	AAK87052	AAK87052 Human imm
35	19	1.0	2535	4	AAK53267	AAK53267 Human pol
36	19	1.0	2565	4	AAK52283	AAK52283 Human pol
37	19	1.0	2674	4	ABLI5134	ABLI5134 Drosophila
38	19	1.0	2832	6	ABN79893	ABN79893 Fungal ZB
39	19	1.0	3579	7	ABT31898	ABT31898 Human bre
40	19	1.0	4172	4	ABLI1397	ABLI1397 Drosophila
41	19	1.0	4357	4	ABLI1338	ABLI1338 Drosophila
42	19	1.0	4365	4	ABLI5066	ABLI5066 Drosophila
43	19	1.0	4387	4	ABLI3810	ABLI3810 Drosophila
44	19	1.0	5742	4	ABLI03491	ABLI03491 Drosophila
45	19	1.0	6804	4	ABLI3824	ABLI3824 Drosophila

#### ALIGNMENTS

RESULT 1  
AAFP25374  
ID AAFP25374 strand; cDNA; 1878 BP.  
XX  
AC AAFP25374;  
XX  
DT 15-MAY-2001 (first entry)  
XX  
DE Nucleotide sequence of the Arabidopsis SGS3 polypeptide.  
XX  
KM SGS3 gene; post-transcriptional inactivation; RNA degradation;  
XX viral resistance; resistance; fatty acid content; protein content; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1878  
FT /tag= a  
FT /product= "SGS3"  
XX  
XX  
PN WO200105951-A2.  
XX  
PD 25-JAN-2001.  
XX  
XX 13-JUL-2000; 2000MO-FR002052.  
PF 16-JUL-1999; 99FR-00009417.  
PR 26-JAN-2000; 2000FR-00001006.  
XX  
PA (AVET ) AVENTIS CROPS/SCIENCE SA.  
PA (INRG ) INST NAT RECH AGRONOMIQUE.  
XX  
XX Beclin C, Elmayan T, Vaucheret H;  
XX  
XX WPI: 2001-159529/16.  
XX P-RSDB; AAB31798.  
XX  
XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
XX resistance in plants and, when inhibited, for increasing transgene  
XX expression.  
XX  
XX Claim 1; Page 32-35; 36pp; French.  
XX  
XX The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide.  
XX The SGS3 gene is essential for post-transcriptional inactivation  
XX (degradation of RNA) and for resistance to viruses. Overexpression of

CC SCS3 results in plants with increased resistance to viruses, while  
CC inactivation of SCS3 in transgenic plants (e.g., by expressing antisense  
CC RNA, by mutation or by homologous recombination) increases the level of  
CC the transgene product. This product may e.g., impart resistance (to  
CC herbicide, insects or pathogens) alter contents of essential fatty acids  
CC or proteins, or is pharmaceutically active, e.g. an immunoglobulin or  
CC interferon

XX Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;  
SQ

Query Match	100.0%;	Score 1878;	DB 4;	Length 1878;
Best Local Similarity	100.0%;	Pred. NO. 0;		
Matches 1878; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

[illegible]

QY	ATTGTGTGGCCCTCCAAATGGTCATCATGATATCTAGACTGGATTAAGGACGATTAACGAT	960
Db	ATTGTGTGGCCCTCCAAATGGTCATCATGATATCTAGACTGGATTAAGGACGATTAACGAT	960
QY	AAATGGCTGTGGCATGGGCAACCAAGATGTGTGAAATCTTTCGCAAGATATGAGGCTCTT	1020
Db	AAATGGCTGTGGCATGGGCAACCAAGATGTGTGAAATCTTTCGCAAGATATGAGGCTCTT	1020
QY	AGAGCAAGCCATTCCTATGCTCCACAGGCGCATGTGTGGATATGATGTTCGATGCTTGAAG	1080
Db	AGAGCAAGCCATTCCTATGCTCCACAGGCGCATGTGTGGATATGATGTTCGATGCTTGAAG	1080
QY	AGCAGTCCCATCTGGGCTATTTGGAGGCCAGCGCTCCACCGGAGTTAGCTTGAGATGGG	1140
Db	AGCAGTCCCATCTGGGCTATTTGGAGGCCAGCGCTCCACCGGAGTTAGCTTGAGATGGG	1140
QY	TTAGATAGAAATTCCTGGGGCTCAAGGCCAGTATGTTTTCTGGAGGGGTTCGCAACTG	1200
Db	TTAGATAGAAATTCCTGGGGCTCAAGGCCAGTATGTTTTCTGGAGGGGTTCGCAACTG	1200
QY	TATGGCTTCCTTGCAAGCAAGATCTGACATATTCATCAACACTCTCAAGGCAAA	1260
Db	TATGGCTTCCTTGCAAGCAAGATCTGACATATTCATCAACACTCTCAAGGCAAA	1260
QY	ACAAAGCTGAAATTCGATTTGAAATCTATACAAAGAGATGGTGTAAAGAGACTGAGCAG	1320
Db	ACAAAGCTGAAATTCGATTTGAAATCTATACAAAGAGATGGTGTAAAGAGACTGAGCAG	1320
QY	ATCTGTGAGCAATATCAGACAGCTGAACTATCTTTAAGAACAGCTCTCAAAACAGAACAG	1380
Db	ATCTGTGAGCAATATCAGACAGCTGAACTATCTTTAAGAACAGCTCTCAAAACAGAACAG	1380
QY	CAGCCCAAGTGTCTTGAGAAATCTCTGAAATTTATGACGAGAAAGCTGTGTGAATCTCA	1440
Db	CAGCCCAAGTGTCTTGAGAAATCTCTGAAATTTATGACGAGAAAGCTGTGTGAATCTCA	1440
QY	GAGATATTCGATGCTGAGACAGAGAACTTAAGTGCATGATGACAGAACAGGGAAAGG	1500
Db	GAGATATTCGATGCTGAGACAGAGAACTTAAGTGCATGATGACAGAACAGGGAAAGG	1500
QY	ATGATGCAACGACAGGTTTTCTATGATTCATTAACAGATTCATGAAAGAAAGAGAC	1560
Db	ATGATGCAACGACAGGTTTTCTATGATTCATTAACAGATTCATGAAAGAAAGAGAC	1560
QY	GAAAGGAGAGAAATTCGAGATGTTGACAGACAGAGAGAGTGCAGAGGTTGTGGCCAG	1620
Db	GAAAGGAGAGAAATTCGAGATGTTGACAGACAGAGAGAGTGCAGAGGTTGTGGCCAG	1620
QY	CAGCAGCAAACTTAAATCCCTCTGACATGACATATTCGCAAGAGAGCTGAGAAAGTG	1680
Db	CAGCAGCAAACTTAAATCCCTCTGACATGACATATTCGCAAGAGAGCTGAGAAAGTG	1680
QY	TCAAGCTTCATGAGTTTCAAGAGAAAGATGAGAGGTTTGTGGAAGAGAGGAGATG	1740
Db	TCAAGCTTCATGAGTTTCAAGAGAAAGATGAGAGGTTTGTGGAAGAGAGGAGATG	1740
QY	CTGATTAAGATCAAGAGAGAGATGGAAGCATGAGAGAGAGGCACTCAGAGAGATTA	1800
Db	CTGATTAAGATCAAGAGAGAGATGGAAGCATGAGAGAGAGGCACTCAGAGAGATTA	1800
QY	TTTGTATCTGAGAAAGATTTGATGAGGCTTTGGAACAGCTCATATTAACAGCATGGCTT	1860
Db	TTTGTATCTGAGAAAGATTTGATGAGGCTTTGGAACAGCTCATATTAACAGCATGGCTT	1860
QY	CAATATGAAGATGATGA 1878	
Db	CAATATGAAGATGATGA 1878	

RESULT 2  
AAAF25373  
ID AAAF25373 standard; DNA; 3275 BP.

XX	AF25373;
AC	
DT	15-MAY-2001 (first entry)
XX	
DE	Genomic sequence of the Arabidopsis SGS3 gene.
XX	
KW	SGS3 gene; post-transcriptional inactivation; RNA degradation;
KW	viral resistance; resistance; fatty acid content; protein content; ss.
XX	
OS	Arabidopsis thaliana.
XX	
EH	Key
FT	Location/Qualifiers
FT	exon
FT	696..1658
FT	/tag= a
FT	1659..1731
FT	/tag= b
FT	1732..2023
FT	/tag= c
FT	2024..2134
FT	/tag= d
FT	2135..2379
FT	/tag= e
FT	2380..2481
FT	/tag= f
FT	2482..2648
FT	/tag= g
FT	2649..2738
FT	/tag= h
FT	2739..2949
FT	/tag= i
XX	
PN	WO200105951-A2.
XX	
PD	25-JAN-2001.
XX	
PE	13-JUL-2000; 2000WO-FR002052.
XX	
PR	16-JUL-1999; 99FR-00009417.
PR	26-JAN-2000; 2000FR-00001006.
XX	
PA	(AVET ) AVENTIS CROSCIENCE SA.
PA	(INRG ) INST NAT RECH AGRONOMIQUE.
XX	
PI	Beclin C, Elmayan T, Vaucheret H;
DR	WPI; 2001-159529/16.
XX	
FT	New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
FT	resistance in plants and, when inhibited, for increasing transgene
FT	expression.
XX	
PS	Claim 1; Page 31-32; 36pp; French.
XX	
CC	The present sequence represents the genomic sequence of the Arabidopsis
CC	thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional
CC	inactivation (degradation of RNA) and for resistance to viruses.
CC	Overexpression of SGS3 results in plants with increased resistance to
CC	viruses, while inactivation of SGS3 in transgenic plants (e.g. by
CC	expressing antisense RNA, by mutation or by homologous recombination)
CC	increases the level of the transgene product. This product may e.g.
CC	impart resistance (to herbicide, insects or pathogens), alter contents of
CC	essential fatty acids or proteins, or is pharmaceutically active, e.g. an
CC	immunoglobulin or interferon
XX	
SO	Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;
XX	
Query Match	51.3%; Score 963; DB 4; Length 3275;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 963; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

Db	696	ATGAGTCTTAGGGCTGGGTCCAAATGTCTTAAGGAAAAAGAACGTTTACGGGTGGTTATAGGCCT	755
QY	61	GAGGTGGAACGTTTGGTTCAAGGTTTGGCAGGGACGACACTGGCTTCTTCAAGAATGAT	120
Db	756	GAGGTGGAACGTTTGGTTCAAGGTTTGGCAGGGACGAGACTGGCTTCTTCAAGAATGAT	815
QY	121	GGAGGAGAGTGGGAGGTCAATTTCCAAAGAACAGAAACAACAGGAAACACTTCTGGA	180
Db	816	GGAGGAGAGTGGGAGGTCAATTTCCAAAGAACAGAAACAACAGGAAACACTTCTGGA	875
QY	181	AAAACTTGGGTTTCTCAGAAATTCGAAATCTTCTAGACCTTGGGGTGGTCAAGCAAGG	240
Db	876	AAAACTTGGGTTTCTCAGAAATTCGAAATCTTCTAGACCTTGGGGTGGTCAAGCAAGG	935
QY	241	AGAGGTAGCAACGATCTGGGAGAGAAACAATGTATCCGGAGAGGTTAACGGCAATGTT	300
Db	936	AGAGGTAGCAACGATCTGGGAGAGAAACAATGTATCCGGAGAGGTTAACGGCAATGTT	995
QY	301	CGGGGCATTCAGCTAACATATCTGGGTGGGGAGAGACGTTTGACAGAAAATATGATTAAC	360
Db	996	CGGGGCATTCAGCTAACATATCTGGGTGGGGAGAGACGTTTGACAGAAAATATGATTAAC	1055
QY	361	AACTTTGGGACCCCAACCTGTATCTCGCCCTCTTGGAAAGAGATGGAATTGGCAG	420
Db	1056	AACTTTGGGACCCCAACCTGTATCTCGCCCTCTTGGAAAGAGATGGAATTGGCAG	1115
QY	421	GCAAGAGAGGTTCTGCTCAGACACACAGCTGTGCAGAGATTCTCTGACGTGAGAGATGAT	480
Db	1116	GCAAGAGAGGTTCTGCTCAGACACACAGCTGTGCAGAGATTCTCTGACGTGAGAGATGAT	1175
QY	481	GTGATATATGTTCTTGAGAGAAAGAAATGATTTCCATGCTTTGGATGATTTCTGATGACGAC	540
Db	1176	GTGATATATGTTCTTGAGAGAAAGAAATGATTTCCATGCTTTGGATGATTTCTGATGACGAC	1235
QY	541	CTTGCAGATGATGATTAATGACTCGAGTGTGAGTCAAAAAGAGCCATGATCCGAAAGCAG	600
Db	1236	CTTGCAGATGATGATTAATGACTCGAGTGTGAGTCAAAAAGAGCCATGATCCGAAAGCAG	1295
QY	601	AATAAGTGTTCAAAAAAGTTCTTTGGCAGAGCTTGGATAGCTTGTGATCGAGCAGATTAAT	660
Db	1296	AATAAGTGTTCAAAAAAGTTCTTTGGCAGAGCTTGGATAGCTTGTGATCGAGCAGATTAAT	1355
QY	661	GAACCAAGAGGCAAGTGGCATTTGCCAGCTTGTCAAAACGGAACCTGGTCCATGATGATTGG	720
Db	1356	GAACCAAGAGGCAAGTGGCATTTGCCAGCTTGTCAAAACGGAACCTGGTCCATGATGATTGG	1415
QY	721	TATTAACCTGCACCCCTTAAGTCAATGAGTCATGCGAGCAAAAAGAGCTAGGCGAGTTAAGTC	780
Db	1416	TATTAACCTGCACCCCTTAAGTCAATGAGTCATGCGAGCAAAAAGAGCTAGGCGAGTTAAGTC	1475
QY	781	CATAGAGAAATGGCTGAAGTTTGAAGAAAGGATCTACAGATGAGAGGCGCATCTGTCAAT	840
Db	1476	CATAGAGAAATGGCTGAAGTTTGAAGAAAGGATCTACAGATGAGAGGCGCATCTGTCAAT	1535
QY	841	CTTTGTGATGATTTATGCGCAGTGAAGGCTTTGGGTGAGATGAAGAAAGATTTATGAA	900
Db	1536	CTTTGTGATGATTTATGCGCAGTGAAGGCTTTGGGTGAGATGAAGAAAGATTTATGAA	1595
QY	901	ATTGTCTGGCCTCCAAATGTCATCATATGAATCTTGAACGCGAATTAAGCAATTAAGAT	960
Db	1596	ATTGTCTGGCCTCCAAATGTCATCATATGAATCTTGAACGCGAATTAAGCAATTAAGAT	1655
QY	961	AAG 963	
Db	1656	AAG 1658	
RESULT 3			
AAF25372/c			
ID AAF25372 standard; DNA: 27 BP.			
AC	AAF25372;		
XX			

DT 15-MAY-2001 (first entry)  
 XX PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.  
 DE  
 XX SGS3 gene; post-transcriptional inactivation; RNA degradation;  
 KM viral resistance; resistance; fatty acid content; protein content;  
 KM PCR primer; ss.  
 OS Arabidopsis thaliana.  
 OS WO200105951-A2.  
 PN 25-JAN-2001.  
 XX 13-JUL-2000; 2000WO-FR002052.  
 XX 16-JUL-1999; 99FR-00009417.  
 PR 26-JAN-2000; 2000FR-00001006.  
 XX (AVET ) AVENTIS CROPS SCIENCE SA.  
 PA (INRG ) INST NAT RECH AGRONOMIQUE.  
 PI Beclin C, Elmayer T, Vaucheret H;  
 PI WPI; 2001-159529/16.  
 DR New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
 XX resistance in plants and, when inhibited, for increasing transgene  
 PT expression.  
 PT Example 1; Page 22; 36pp; French.  
 XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis  
 CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-  
 CC transcriptional inactivation (degradation of RNA) and for resistance to  
 CC viruses. Overexpression of SGS3 results in plants with increased  
 CC resistance to viruses, while inactivation of SGS3 in transgenic plants  
 CC (e.g. by expressing antisense RNA, by mutation or by homologous  
 CC recombination) increases the level of the transgene product. This product  
 CC may e.g. impart resistance (to herbicide, insects or pathogens), alter  
 CC contents of essential fatty acids or proteins, or is pharmaceutically  
 CC active, e.g. an immunoglobulin or interferon  
 CC  
 SQ Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;  
 Query Match 1.3%; Score 24; DB 4; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 0.71;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1855 GGCTTCACATGAGATGATGTA 1878  
 Db 27 GGCTTCACATGAGATGATGTA 4  
 RESULT 4  
 AAF25371  
 ID AAF25371 standard; DNA; 23 BP.  
 AC AAF25371;  
 XX 15-MAY-2001 (first entry)  
 DT PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.  
 DE  
 XX SGS3 gene; post-transcriptional inactivation; RNA degradation;  
 KM viral resistance; resistance; fatty acid content; protein content;  
 KM PCR primer; ss.  
 OS Arabidopsis thaliana.  
 OS WO200105951-A2.  
 PN 25-JAN-2001.  
 XX

XX 13-JUL-2000; 2000WO-FR002052.  
 PF 16-JUL-1999; 99FR-00009417.  
 XX 26-JAN-2000; 2000FR-00001006.  
 PR (AVET ) AVENTIS CROPS SCIENCE SA.  
 PA (INRG ) INST NAT RECH AGRONOMIQUE.  
 XX Beclin C, Elmayer T, Vaucheret H;  
 PI WPI; 2001-159529/16.  
 DR New SGS3 gene from Arabidopsis thaliana, useful for increasing virus  
 XX resistance in plants and, when inhibited, for increasing transgene  
 PT expression.  
 PT Example 1; Page 22; 36pp; French.  
 XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis  
 CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-  
 CC transcriptional inactivation (degradation of RNA) and for resistance to  
 CC viruses. Overexpression of SGS3 results in plants with increased  
 CC resistance to viruses, while inactivation of SGS3 in transgenic plants  
 CC (e.g. by expressing antisense RNA, by mutation or by homologous  
 CC recombination) increases the level of the transgene product. This product  
 CC may e.g. impart resistance (to herbicide, insects or pathogens), alter  
 CC contents of essential fatty acids or proteins, or is pharmaceutically  
 CC active, e.g. an immunoglobulin or interferon  
 CC  
 SQ Sequence 23 BP; 6 A; 4 C; 7 G; 6 T; 0 U; 0 Other;  
 Query Match 1.1%; Score 20; DB 4; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 ATGAGTTCTAGGCTGCTCC 20  
 Db 4 ATGAGTTCTAGGCTGCTCC 23  
 RESULT 5  
 AAF12773  
 ID AAF12773 standard; cDNA; 596 BP.  
 AC AAF12773;  
 XX 13-MAR-2001 (first entry)  
 DT Aspergillus oryzae EST SEQ ID NO:5296.  
 DE  
 XX Multiple gene expression; filamentous fungal cell; EST;  
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KM culture condition; environmental stress; spore morphogenesis;  
 KM metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX Aspergillus oryzae.  
 OS Aspergillus oryzae.  
 OS WO200056762-A2.  
 PN 28-SEP-2000.  
 XX 22-MAR-2000; 2000WO-US007781.  
 PR 22-MAR-1999; 99US-00273623.  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 PI Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;  
 PI WPI; 2000-594572/56.  
 DR

XX Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags.  
XX  
PS Claim 88; Page 2210; 3161pp; English.  
XX  
CC The present invention describes a method for monitoring differential  
CC expression of genes in a first filamentous fungal (FF) cell relative to  
CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring the  
CC global expression of genes from FF cells allows the production potential  
CC of the microorganisms to be improved. New genes may be discovered,  
CC possible functions of unknown open reading frames can be identified and  
CC gene copy number variation and stability can be monitored. The expression  
CC of genes can be used to study how FF cells adapt to changes in culture  
CC conditions, environmental stress, spore morphogenesis, recombination,  
CC metabolic or catabolic pathway engineering. Using ESTs provides several  
CC advantages over genomic or random cDNA clones including elimination of  
CC redundancy as one spot on an array equals one gene or open reading frame,  
CC and organization of the microarrays based on function of the gene  
CC products to facilitate analysis of the results. AAF07478 to AAF11247  
CC represents ESTs from *Fusarium venenatum*; AAF11248 to AAF11853 represents  
CC ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from  
CC *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from  
CC *Trichoderma reesei*, which are all specifically claimed in the present  
CC invention  
XX  
SQ Sequence 596 BP; 144 A; 168 C; 150 G; 134 T; 0 U; 0 Other;  
XX  
Query Match 1.1%; Score 20; DB 3; Length 596;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 142 TTCCAGAGAGAACAGAACAA 161  
Db 431 TTCCAGAGAGAGAACAGAACAA 450  
XX  
RESULT 6  
AAH93017/c  
ID AAH93017 standard; DNA; 700 BP.  
XX  
AC AAH93017;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Human inflammatory bowel disease related gene fragment IGR3310a.  
XX  
KW Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
KW single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;  
XX chromosome 5q31-33; forensic test; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200142511-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 11-DEC-2000; 2000WO-US033632.  
XX  
PR 10-DEC-1999; 99US-0170257P.  
PR 10-APR-2000; 2000US-0196046P.  
XX  
XX (WHED) WHITEHEAD INST BIOMEDICAL RES.  
PA (ELLT-) ELLIPISIS BIOTHERAPEUTICS CORP.  
XX  
XX Daly M, Hudson TJ, Lander ES, Rioux J, Simionovitch K;  
XX

DR WPI; 2001-367874/38.  
XX  
XX Testing for the presence of polymorphisms associated with inflammatory  
PT bowel disease, using a hybridization assay.  
PT  
XX  
XX Disclosure; Page 381; 463pp; English.  
XX  
XX The present invention describes a method for detecting the presence of  
CC polymorphisms associated with inflammatory bowel diseases such as  
CC ulcerative colitis and Crohn's disease. The methods can be used to detect  
CC the presence of genetic polymorphisms associated with inflammatory bowel  
CC disease and correlating their occurrence with disease states. They may be  
CC used in this way for phenotypic correlations, forensics, paternity  
CC testing, medicine and genetic analysis. The present sequence is a gene  
CC containing a polymorphic site described in the exemplification of the  
CC invention  
XX  
SQ Sequence 700 BP; 156 A; 158 C; 149 G; 237 T; 0 U; 0 Other;  
XX  
Query Match 1.1%; Score 20; DB 4; Length 700;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1696 TTTCAGAGAGAGAGATGGA 1715  
Db 415 TTTCAGAGAGAGAGATGGA 396  
XX  
RESULT 7  
AAA81465/c  
ID AAA81465 standard; DNA; 47475 BP.  
XX  
AC AAA81465;  
XX  
DT 04-DEC-2000 (first entry)  
XX  
DE N. meningitidis partial DNA sequence gnm\_13 SEQ ID NO:13.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
XX Neisseria meningitidis; B; MenB; ds.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO2000022430-A2.  
XX  
PD 20-APR-2000.  
XX  
PF 08-OCT-1999; 99WO-US023573.  
XX  
PR 09-OCT-1998; 98US-0103794P.  
PR 30-APR-1999; 99US-0132068P.  
XX  
XX (CHIR) CHIRON CORP.  
XX  
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,  
PI Masiangni V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V,  
PI Rappuoli R, Pizza M;  
XX  
XX WPI; 2000-318079/27.  
XX  
XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be used  
PT in the diagnosis and treatment of *N. meningitidis* infection and other  
PT Neisserial infections; for example, *N. gonorrhoea*.  
XX  
XX Claim 7; Page 383-397; 1760pp; English.  
XX  
XX The present invention describes methods of obtaining immunogenic proteins  
CC from *Neisseria* genomic sequences. AAA81453 to AAA82414 represent  
CC specifically claimed *Neisseria meningitidis* genomic DNA sequences;  
CC AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA  
CC sequences and their corresponding proteins; AAA81254 to AAA81259 and  
CC AAA81304 to AAA81321 represent PCR primers used in the isolation of



CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent  
 CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all  
 CC used in the exemplification of the present invention. The nucleic acid  
 CC sequences, protein sequences, and antibodies against them, can be used in  
 CC the manufacture of a composition. The composition can be used as a  
 CC medicament (or in the manufacture of a medicament) for treating,  
 CC preventing or diagnosing infection due to Neisserial bacteria. For  
 CC example, some of the identified proteins could be components of vaccines  
 CC against Meningococcus B; against all serotypes; and/or against all  
 CC pathogenic Neisseriae. Identification of sequences from the bacterium  
 CC will also facilitate production of biological probes, particularly  
 CC organism-specific probes. Attempts to make efficacious Meningococcus B  
 CC vaccines have failed mainly due to antigen tolerance. Multivalent  
 CC vaccines have also been tried but none have successfully overcome  
 CC antigenic variability. The provision of further, complete sequences may  
 CC provide an opportunity to identify secreted or surface exposed proteins  
 CC that may be presumed targets for the immune system and which are not  
 CC antigenically variable or at least more conserved than other more  
 CC variable regions

XX SQ Sequence 47475 BP; 11600 A; 13758 C; 12290 G; 9826 T; 0 U; 1 Other;  
 XX Query Match 1.1%; Score 20; DB 3; Length 47475;  
 XX Best Local Similarity 100.0%; Pred. No. 55;  
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 GAACAGTTGTTCAAGTTT 86  
 DB 23354 GAACAGTTGTTCAAGTTT 23335

RESULT 8  
 AAA81489\_5/c  
 Continuation (6 of 9) of AAA81489 from base 500001 (N. meningitidis partial DNA sequence  
 WP Sequence split into 9 fragments LOCUS AAA81489 Accession Aaa81489  
 WP Fragment Name Begin End  
 WP AAA81489\_0 1 110000  
 WP AAA81489\_1 100001 210000  
 WP AAA81489\_2 200001 310000  
 WP AAA81489\_3 300001 410000  
 WP AAA81489\_4 400001 510000  
 WP AAA81489\_5 500001 610000  
 WP AAA81489\_6 600001 710000  
 WP AAA81489\_7 700001 810000  
 WP AAA81489\_8 800001 837096

Query Match 1.1%; Score 20; DB 3; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 GAACAGTTGTTCAAGTTT 86  
 DB 32595 GAACAGTTGTTCAAGTTT 32576

RESULT 9  
 AB088198/c  
 ID AB088198 standard; cDNA; 154902 BP.  
 XX AC AB088198;  
 XX DT 18-SEP-2002 (first entry)  
 XX DE Human osteoblast differentiation related cDNA SEQ ID NO 105.  
 XX KW Human; osteoblast; stem cell differentiation; bone tissue deposition;  
 KM osteoporosis; osteopathic; ss.  
 OS Homo sapiens.  
 XX WO200250301-A2.  
 XX PN 27-JUN-2002.  
 XX PD

XX PF 18-DEC-2001; 2001WO-US048276.  
 XX PR 18-DEC-2000; 2000US-0255882P.  
 XX PR 24-APR-2001; 2001US-0285691P.  
 XX PA (GENE-) GENE LOGIC INC.  
 XX PA (PROC) PROCTER & GAMBLE CO.  
 XX PI J D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;  
 XX PI Mertz L;  
 XX DR WPI; 2002-557663/59.  
 XX PT Use of genes and their expression profiles associated with osteoblast  
 XX PT differentiation for screening modulators bone formation, for diagnosing  
 XX PT or treating e.g. osteoporosis, or as markers for the differentiation  
 XX PT process.

XX SQ Claim 1; SEQ ID NO 105; 78bp + Sequence Listing; English.

XX CC The invention relates to genes and their expression profiles are used  
 XX CC for: (a) screening modulators of precursor stem cell differentiation into  
 XX CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal  
 XX CC deposition of bone tissue, abnormal rate of osteoblast formation or  
 XX CC osteoporosis; or (c) treating or monitoring treatment of the conditions  
 XX CC cited in (b), or monitoring the progression of bone tissue deposition;  
 XX CC specific conditions include postmenopausal osteoporosis, glucocorticoid  
 XX CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-  
 XX CC induced abnormalities in bone formation or bone loss, conditions that  
 XX CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),  
 XX CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome  
 XX CC or fibrous dysplasia. The present sequence is that of an osteoblast  
 XX CC differentiation associated cDNA marker of the invention. Note: The  
 XX CC sequence data for this patent did not form part of the printed  
 XX CC specification, but was obtained in electronic format directly from WIPO  
 XX CC at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 154902 BP; 43917 A; 31458 C; 32848 G; 46679 T; 0 U; 0 Other;  
 XX Query Match 1.1%; Score 20; DB 6; Length 154902;  
 XX Best Local Similarity 100.0%; Pred. No. 54;  
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1712 TGGAGAGTTGTGGAAGG 1731  
 DB 128153 TGGAGAGTTGTGGAAGG 128134

RESULT 10  
 AAF21612/c  
 ID AAF21612 standard; DNA; 349980 BP.  
 XX AC AAF21612;  
 XX DT 13-MAR-2001 (first entry)  
 XX DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.  
 XX KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic vaccine;  
 XX KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 XX KW ds.  
 XX OS Neisseria meningitidis.  
 XX WO20006791-A1.  
 XX PN 09-NOV-2000.  
 XX PD 08-MAR-2000; 2000WO-US005928.  
 XX PF 30-APR-1999; 99US-0132068P.  
 XX PR 08-OCT-1999; 99WO-US023573.  
 XX PR

PR 28-FEB-2000; 2000GB-00004695.  
 XX (CHIR) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX  
 PI Pizsa M, Hickey E, Peterson J, Tettelin H, Venter JC, Scarlato V;  
 PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Sgarbiato V;  
 PI Rappelli R, Frazer CM, Grandi G;  
 DR WPI; 2000-647603/52.  
 XX  
 PT Neisseria meningitidis B full length genome sequence and open reading  
 PT frames are used to detect, treat and prevent Neisserial infections.  
 XX  
 PS Claim 7; Appendix A; 692pp; English.  
 XX  
 CC The present invention describes the full length genome of Neisseria  
 CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613  
 CC represent fragments of the NMB genomic sequence, as the sequence was too  
 CC long to go in a record on its own it was split into 8 sequences which  
 CC overlap each other at the beginning and end of each sequence by 49980 bp  
 CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of  
 CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of  
 CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins  
 CC given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR  
 CC primers which are used in the exemplification of the present invention.  
 CC The NMB genome and fragments from it have antibacterial activity, and can  
 CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins  
 CC and/or antibodies which binds to the proteins can be used in compositions  
 CC for treating or preventing infection due to Neisserial bacteria or as a  
 CC diagnostic reagent for detecting the presence of Neisserial bacteria or  
 CC of antibodies raised to Neisserial bacteria. Computers, computer memory,  
 CC computer storage medium or computer databases can be used in a search to  
 CC identify open reading frames (ORFs) or coding sequences within the NMB  
 CC genome. The DNA sequences provide further opportunities to find antigenic  
 CC or immunogenic proteins which are more effective in vaccines than the  
 CC outer membrane proteins currently used  
 XX  
 SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 U; 0 Other;  
 Query Match 1.1%; Score 20; DB 3; Length 349980;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 67 GAACAGTGTGTTCAAGTTT 86  
 Db 167824 GAACAGTGTGTTCAAGTTT 167805  
 RESULT 11  
 ID AAA68016/c  
 ID AAA68016 standard; DNA; 481 BP.  
 AC AAA68016;  
 XX  
 DT 24-OCT-2000 (first entry)  
 XX  
 DE Eucalyptus grandis CCR nucleotide sequence SEQ ID NO:109.  
 XX  
 KW Plant; lignin; lignin biosynthetic pathway; Eucalyptus grandis;  
 KW Pinus radiata; Monterey pine; ds.  
 XX  
 OS Eucalyptus grandis.  
 XX  
 PN WO200022099-A1.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 06-OCT-1999; 99WO-NZ000168.  
 XX  
 PR 09-OCT-1998; 98US-00168789.  
 PR 14-UTL-1999; 99US-0143811P.  
 XX

PA (GENE-) GENESIS RES & DEV CORP LTD.  
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX  
 XX  
 PI Blokeberg LN, Hawukkala IU;  
 DR WPI; 2000-317962/27.  
 XX  
 XX  
 PT Novel polynucleotide encoding enzymes involved in lignin-biosynthetic  
 PT pathway useful for producing transgenic plants especially eucalyptus and  
 PT pine species having altered lignin content, composition and structure.  
 XX  
 PS Claim 1; Page 91-92; 213pp; English.  
 XX  
 CC The present invention describes isolated polynucleotides and proteins  
 CC encoding and representing the enzymes cinnamate 4-hydroxylase (CAH),  
 CC coumarate 3-hydroxylase (C3H), phenolase (PN), O-methyl transferase  
 CC (OMT), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl-CoA reductase  
 CC (CCR), phenylalanine ammonia-lyase (PAL), 4-coumarate:CoA ligase (4CL),  
 CC coniferol glucosyl transferase (CGT), coniferin beta-glucosidase (CBG),  
 CC laccase, peroxidase, ferulate-5-hydroxylase (F5H), alpha-amylase, caffeic  
 CC acid methyl transferase, caffeoyl CoA methyl transferase, coumarate CoA  
 CC ligase, cyclochrome P450 1LX1A, diphenol oxidase, flavanol glucosyl  
 CC transferase, flavonoid hydroxylase, and isoflavone reductase, which are  
 CC involved in the lignin biosynthetic pathway. The polynucleotides can be  
 CC used for modulating lignin content, lignin composition and the structure  
 CC of a plant, especially eucalyptus and pine species, and for modifying the  
 CC activity of an enzyme involved in lignin biosynthetic pathway, and for  
 CC producing a plant having altered lignin content, composition and  
 CC structure. They can be used for designing probes and primers useful for  
 CC detecting similar DNA and RNA sequences in any organism and for PCR  
 CC amplification. The lignin content can be efficiently modified using the  
 CC polynucleotides. AAB57908 to AAB68201 and AAB16341 to AAB16449 represent  
 CC polynucleotide and protein sequences used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 481 BP; 122 A; 118 C; 111 G; 130 T; 0 U; 0 Other;  
 Query Match 1.0%; Score 19; DB 3; Length 481;  
 Best Local Similarity 100.0%; Pred. No. 1,8e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 74 TGGTTCAGGTTTGGCAGG 92  
 Db 470 TGGTTCAGGTTTGGCAGG 452  
 RESULT 12  
 ID ADD41766/c  
 ID ADD41766 standard; DNA; 481 BP.  
 AC ADD41766;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Cinnamoyl-CoA reductase DNA #21.  
 XX  
 KW ds; plant; lignin content; lignin composition; lignin structure;  
 KW lignin biosynthesis; gene.  
 XX  
 OS Eucalyptus grandis.  
 XX  
 PN US2003131373-A1.  
 XX  
 PD 10-UTL-2003.  
 XX  
 PF 18-UTN-2002; 2002US-00174693.  
 XX  
 PR 11-SEP-1996; 96US-00713000.  
 PR 21-NOV-1997; 97US-00975316.  
 PR 09-OCT-1998; 98US-00169789.  
 PR 14-UTL-1999; 99US-0143813P.  
 PR 12-UTL-2000; 2000US-00615192.  
 XX



PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249246P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 06-DEC-2000; 2000US-0256719P.  
 PR 08-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251899P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-483426/52.  
 XX  
 PT P-PSDB; AAM91465.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 XX useful for preventing, diagnosing and/or treating cancers and metastasis.  
 PS Claim 1; SEQ ID NO 9306; 3071bp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169

CC represent sequences used in the exemplification of the present invention  
 XX  
 SQ Sequence 484 BP; 179 A; 66 C; 106 G; 131 T; 0 U; 2 Other;  
 Query Match 1.0%; Score 19; DB 4; Length 484;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1767 GGAAGACATGAGAGAGG 1785  
 Db 323 GGAAGACATGAGAGAGG 341  
 RESULT 14  
 AAV23847/C  
 ID AAV23847 standard; DNA; 495 BP.  
 XX  
 AC AAV23847;  
 XX  
 DT 31-JUL-1998 (first entry)  
 XX  
 DE Plant CCR enzyme DNA sequence.  
 XX  
 KW Lignin biosynthetic pathway; eucalyptus; pine; transgenic plant;  
 KM Lignin content; tree processing; cellulose fibre; ss.  
 XX  
 OS Eucalyptus grandis.  
 XX  
 FN WO9811205-A2.  
 XX  
 PD 19-MAR-1998.  
 XX  
 PF 10-SEP-1997; 97WO-NZ000112.  
 XX  
 PR 11-SEP-1996; 96US-00713000.  
 XX  
 PA (GENE-) GENESIS RES & DEV CORP LTD.  
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
 XX  
 PI Blokberg LN, Grierson AW, Havukkala IU;  
 XX  
 DR WPI; 1998-207374/18.  
 XX  
 PT Sequences useful for modification of plant lignin content or structure -  
 PT from Eucalyptus grandis (eucalyptus) and Pinus radiata (pine) are  
 PT associated with lignin biosynthesis pathway, useful e.g. in paper  
 PT industry.  
 PS Claim 1; Page 39; 82pp; English.  
 XX  
 CC This sequence represents a fragment of the CCR enzyme coding sequence. It  
 CC is an example of a DNA sequence of the invention, which are from  
 CC Eucalyptus grandis (eucalyptus) and Pinus radiata (pine) associated with  
 CC the lignin biosynthesis pathway. Constructs containing the DNA sequences  
 CC can be used to produce transgenic plants or plant cells, especially woody  
 CC plants e.g. eucalyptus or pine species but also e.g. monocotyledons or  
 CC dicotyledons; by stably incorporating the constructs into the plant  
 CC genome. The lignin content or structure, or activity of a specific enzyme  
 CC in the plant, can therefore be modulated. Reductions in lignin content or  
 CC changes in composition are useful in tree processing for paper. High  
 CC lignin content results in energy- and chemical-intensive separation  
 CC methods in order to obtain the pure cellulose fibre required. Reductions  
 CC in lignin content may also be useful for forage crops, whilst increases  
 CC or changes in composition may be desirable to increase the mechanical  
 CC strength of wood, change its colour or increase its resistance to rot.  
 CC The sequences are also useful as probes to isolate DNA sequences encoding  
 CC enzymes involved in the lignin biosynthesis pathway from other plant  
 CC species  
 XX  
 SQ Sequence 495 BP; 126 A; 121 C; 116 G; 132 T; 0 U; 0 Other;  
 Query Match 1.0%; Score 19; DB 2; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 TGGTTCAAGTTTGGCAGG 92  
 |||||  
 DB 484 TGGTTCAAGTTTGGCAGG 466

Search completed: April 9, 2004, 02:09:06  
 Job time : 482.978 secs

RESULT 15  
 AAZ06848/C  
 ID AAZ06848 standard; cDNA, 495 BP.

XX AAZ06848;  
 XX

DT 09-NOV-1999 (first entry)

DE Eucalyptus cinamoyl-CoA reductase (CCR) partial cDNA 2.

XX Lignin; biosynthesis; forage crop; wood; paper production;  
 KM transgenic plant; ss.  
 XX

OS Eucalyptus grandis.

XX US952486-A.

PN 14-SEP-1999.

PD 21-NOV-1997; 97US-00975316.

XX 11-SEP-1996; 96US-00713000.

PR (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Grierson AW, Bloksberg LN, Havukkala I;

PI WPI; 1999-527029/44.

DR Isolated DNA sequence encoding enzymes from the lignin synthetic pathway  
 XX useful for generating plants with an altered lignin content.  
 PT

XX Example 1; Col 37-38; 48pp; English.

XX This sequence represents a cinamoyl-CoA reductase (CCR) partial cDNA  
 CC from Eucalyptus grandis. This enzyme is involved in the biosynthesis of  
 CC lignin, an insoluble polymer which is primarily responsible for the  
 CC rigidity of plant stems. Lignin serves as a matrix around the  
 CC polysaccharide components of some plant cell walls. The higher the lignin  
 CC content, the more rigid the plant. Lignin also plays a role in disease  
 CC resistance of plants by impeding the penetration and propagation of  
 CC pathogenic agents. Lignin is formed by polymerisation of at least three  
 CC different monolignols (para-coumaryl alcohol, coniferyl alcohol and  
 CC sinapyl alcohol). These three monolignols are synthesised by similar  
 CC pathways from phenylalanine in a multistep process and are believed to be  
 CC polymerised into lignin via a free radical mechanism. The lignin content  
 CC of plants can be altered using DNA sequences encoding these enzymes.  
 CC Lignin content can be increased by incorporation of additional copies of  
 CC genes encoding these enzymes into the target plant. This could be  
 CC beneficial for increasing the mechanical strength of wood. Similarly, a  
 CC decrease in lignin content can be obtained by transforming the target  
 CC plant with antisense copies of such genes. This may be beneficial in  
 CC plants used as forage crops for livestock (lignin is indigestible) and in  
 CC trees used in paper manufacture  
 CC

SQ Sequence 495 BP; 126 A; 121 C; 116 G; 132 T; 0 U; 0 Other;

Query Match 1.0%; Score 19; DB 2; Length 495;

Best local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 TGGTTCAAGTTTGGCAGG 92  
 |||||  
 DB 484 TGGTTCAAGTTTGGCAGG 466

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2004, 20:08:06 ; Search time 3159.03 Seconds  
(without alignments)  
1752.632 Million cell updates/sec

Title: US-10-030-829-2  
Perfect score: 1878  
Sequence: 1 atgagttcctcag9ctgtctcc.....ttccaatgaagtatgta 1878

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :  
EST:  
1: em\_estda:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	565	30.1	693	28	AQ959658 LEREM21TF
2	534	28.4	606	28	AQ959657 LEREM21TF
3	513	27.3	619	9	AV566465 AV566465
4	512	27.3	647	9	AU236368 AU236368

Result No.	Score	Query Match	Length	DB ID	Description
5	474	25.2	613	28	AQ964580 LEREM21TF
6	339	17.5	380	29	AV525508 AV525508
7	295	15.7	457	29	CNS00029H Arabidops
8	273	14.5	523	28	AQ959659 LEREM21TF
9	244	13.0	421	9	AU227277 AU227277
10	228	12.1	290	28	AQ010650 F27C8TRC
11	169	9.0	548	9	AI999551 AI999551
12	148	8.1	197	28	BH169457 SALK 0013
13	153	7.9	443	28	BH168393 SALK 0013
14	148	7.9	443	28	BH169448 SALK 0013
15	47	2.5	232	28	AQ964581 LEREM21TF
16	38	2.0	384	29	BK653772 Arabidops
17	38	2.0	384	29	BK653772 Arabidops
18	29	1.5	644	14	CD835509 BN45.045L
19	29	1.5	697	28	BH535855 BOHXR13TF
20	28	1.5	587	13	BO507998 EST615413
21	24	1.3	123	29	BK654916 Arabidops
22	24	1.3	405	29	BK287411 Arabidops
23	22	1.2	270	13	BY023513 BY023513
24	22	1.2	562	12	BM62455 daq1b10.
25	22	1.2	621	12	BJ625937 BJ625937
26	22	1.2	675	14	CA798850 EST1648 M
27	22	1.2	708	14	CD815300 BN15.0250
28	22	1.2	896	29	CG969741 MBERH11TR
29	22	1.2	967	9	AV254322 AV254322
30	22	1.2	2684	11	AK029504 Mus muscu
31	22	1.1	258	10	BB578249 BB578249
32	21	1.1	319	12	BG279367 b2f06np.f
33	21	1.1	350	14	CF598325 NCEST3d89
34	21	1.1	376	14	BG543396 E0910 Ch1
35	21	1.1	396	14	CB831627 NCEST3a63
36	21	1.1	402	10	BF248666 NCEST3a01
37	21	1.1	409	14	CF371330 NCEST3d50
38	21	1.1	414	10	BR824285 NCEST3a24
39	21	1.1	419	10	BF249331 NCEST3a09
40	21	1.1	424	10	BF716900 NCEST3a18
41	21	1.1	425	14	CF260387 NCEST3d04
42	21	1.1	432	10	BF716619 NCEST3a15
43	21	1.1	432	10	BF249140 NCEST3a07
44	21	1.1	433	12	BG235108 NCEST3a40
45	21	1.1	434	10	BF248638 NCEST3a01

## ALIGNMENTS

RESULT 1  
LOCUS AQ959658/c  
DEFINITION LEREM21TF LERE Arabidopsis thaliana genomic clone LEREM21, genomic survey sequence.  
ACCESSION AQ959658  
VERSION AQ959658.1 GI:6787359  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana

REFERENCE  
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eustosids II; Brassicales; Brassicaceae; Arabidops.  
1 (bases 1 to 693)  
Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utebach, T., Feldblyum, T., Liang, F., Creasy, T. and Fraser, C.M.  
Genomic survey sequencing of landberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms  
Unpublished (2000)  
Contact: Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: at@igir.org

## JOURNAL COMMENT

For additional information, see <http://www.tigr.org/cdb/at.html>

Seq primer: TF  
Class: Shotgun.

FEATURES  
source  
Location/Qualifiers

1..693  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="LANDSBERG\_ERECLA"  
/db\_xref="taxon:3702"  
/clone\_lib="LEREM21"  
/note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was sheared to 0.6-0.8 Kbp before ligation."

## ORIGIN

Query Match 30.1%; Score 565; DB 28; Length 693;  
Best Local Similarity 99.8%; Pred. No. 1.6e-276;  
Matches 615; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 134 AGGTCAATTTCCAGAGAACAGAACAAACAGAAACACTTCTGGAAAAAATTGGGTTT 193
DB 688 AGGTCAATTTCCAGAGAACAGAACAAACAGAAACACTTCTGGAAAAAATTGGGTTT 629
QY 194 CTCAGAAATTCGAATCTCTTGAAGCTTGGGGTGTGACGACAAAGGAGAGATGACAG 253
DB 628 CTCAGAAATTCGAATCTCTTGAAGCTTGGGGTGTGACGACAAAGGAGAGATGACAG 569
QY 254 TATCTGGAGAGAGAGAAACATGTATCCGGAGAGATTAACGCAATGGTCGGGCAATTCAG 313
DB 568 TATCTGGAGAGAGAGAAACATGTATCCGGAGAGATTAACGCAATGGTCGGGCAATTCAG 509
QY 314 CTAAATATCTGTGTGGGAGAGAGCTTGAAGCAAAAGTATGATTAACACTTTGTGGC 373
DB 508 CTAAATATCTGTGTGGGAGAGAGCTTGAAGCAAAAGTATGATTAACACTTTGTGGC 449
QY 374 CCCCACTGTATCTGGGCTCTTCTTGAAGAGAGATGAAATGGCAGGCAAGAGAGTT 433
DB 448 CCCCACTGTATCTGGGCTCTTCTTGAAGAGAGATGAAATGGCAGGCAAGAGAGTT 389
QY 434 CTGCTCAGACACAGCTGTGAGAGAGTTCTGACGTTGAGAGATGATGTGATTAATCTT 493
DB 388 CTGCTCAGACACAGCTGTGAGAGAGTTCTGACGTTGAGAGATGATGTGATTAATCTT 329
QY 494 CTGAGAGAGAGATGATTCGATCTTGGATGATGATGATGATGATGATGATGATGATG 553
DB 328 CTGAGAGAGAGATGATTCGATCTTGGATGATGATGATGATGATGATGATGATGATG 269
QY 554 ATTATGATCTGGAGTGTGATCAAAAGAGCATGATGATGATGATGATGATGATGATGATG 613
DB 268 ATTATGATCTGGAGTGTGATCAAAAGAGCATGATGATGATGATGATGATGATGATGATG 209
QY 614 AAAAGTTCTTGGAGCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATG 673
DB 208 AAAAGTTCTTGGAGCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATG 149
QY 674 AGTGGCATTTGTCCACTTGTGAGAGAGAGCTGGATGATGATGATGATGATGATGATGATG 733
DB 148 AGTGGCATTTGTCCACTTGTGAGAGAGAGCTGGATGATGATGATGATGATGATGATGATG 89
QY 734 CTCTACTAGCTCATGC 749
DB 88 CTCTACTAGCTCATGC 73

```

RESULT 2  
AQ959657/c 606 bp DNA linear GSS 28-JAN-2000  
LOCUS LEREM21TF LERE Arabidopsis thaliana genomic clone LEREM21, genomic  
DEFINITION survey sequence.  
ACCESSION AQ959657  
VERSION AQ959657.1 GI:6787358  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
COMMENT

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 606)  
Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utebach, T.,  
Feldblyum, T., Liang, F., Creasy, T., and Fraser, C.M.  
Genomic survey sequencing of Landsberg erecta ecotype of  
Arabidopsis thaliana and identification of sequence-based  
polymorphisms  
Unpublished (2000)  
Contact: Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 0208  
Email: atc@tigr.org  
For additional information, see <http://www.tigr.org/db/at/at.html>  
Seq primer: TF  
Class: Shotgun.

FEATURES  
source

1..606  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="LANDSBERG\_ERECLA"  
/db\_xref="taxon:3702"  
/clone\_lib="LEREM21"  
/note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was sheared to 0.6-0.8 Kbp before ligation."

## ORIGIN

Query Match 28.4%; Score 534; DB 28; Length 606;  
Best Local Similarity 100.0%; Pred. No. 1e-260;  
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 216 ACCTTGGGGTGTGACGACGACAGAGAGAGTACAACTATCTGGAGAGAGAAACAATGT 275
DB 606 ACCTTGGGGTGTGACGACGACAGAGAGAGTACAACTATCTGGAGAGAGAAACAATGT 547
QY 276 ATCCGGAGAGAGTAAACGCAATGTGCGGGCACTTCAACTAATATCTGTCGGGAGC 335
DB 546 ATCCGGAGAGAGTAAACGCAATGTGCGGGCACTTCAACTAATATCTGTCGGGAGC 487
QY 336 AGCGTTGACAGAAATGATGATTAACAACTTGTGACACCCCACTGATATCTGCGCCCTCC 395
DB 486 AGCGTTGACAGAAATGATGATTAACAACTTGTGACACCCCACTGATATCTGCGCCCTCC 427
QY 396 TTGGAAGAGATGGAATTTGGCAGAGAGAGAGTTCTGCTCAGACACAGCTGTGCA 455
DB 426 TTGGAAGAGATGGAATTTGGCAGAGAGAGAGTTCTGCTCAGACACAGCTGTGCA 367
QY 456 GGAATTTCTGACGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 515
DB 366 GGAATTTCTGACGTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 307
QY 516 TGTCTTGATGATTTGATGACGACCTTCAAGTATGATTTGATGATGATGATGATGATG 575
DB 306 TGTCTTGATGATTTGATGACGACCTTCAAGTATGATTTGATGATGATGATGATGATG 247
QY 576 AAAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 635
DB 246 AAAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 187
QY 636 TAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 695
DB 186 TAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 127
QY 696 GAACGACCTGTGTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 749
DB 126 GAACGACCTGTGTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 73

```

RESULT 3

AV566465/c  
 LOCUS AV566465 619 bp mRNA linear EST 07-SEP-2000  
 DEFINITION AV566465 Arabidopsis thaliana green siliques Columbia Arabidopsis  
 thaliana cDNA clone SQ244b06f 3', mRNA sequence.  
 ACCESSION AV566465  
 VERSION AV566465.1 GI:8737917  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 619)  
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries  
 DNA Res. 7 (3), 175-180 (2000)  
 JOURNAL MEDLINE  
 PUBMED 20363093  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/  
 location/Qualifiers  
 1. 619  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="SQ244b06f"  
 /issue\_type="green siliques"  
 /clone\_lib="Arabidopsis thaliana green siliques Columbia"  
 /note="Vector: pBluescriptII SK-, Site\_1: EcoRI; Site\_2:  
 XhoI"

ORIGIN  
 Query Match 27.3%; Score 513; DB 9; Length 619;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-250;  
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1291 CAAGAGTGGTGTAAAGAGCTAGGACAGATCTGTGAGACAATCAGACGTGAATAC 1350  
 DB 619 CAAGAGTGGTGTAAAGAGCTAGGACAGATCTGTGAGACAATCAGACGTGAATAC 560  
 QY 1351 TTTAAGAACAGCTCTCAAAACAGAACACAGCCCAAGGTGCTTGAAGATCTTGAA 1410  
 DB 559 TTTAAGAACAGCTCTCAAAACAGAACACAGCCCAAGGTGCTTGAAGATCTTGAA 500  
 QY 1411 ATTATAGGAGAAAGCTGCTTAAGACTGACAGAGATTAATGCGATCGTGAACAGAACT 1470  
 DB 499 ATTATAGGAGAAAGCTGCTTAAGACTGACAGAGATTAATGCGATCGTGAACAGAACT 440  
 QY 1471 AAGATCAGCATAAACAGAACAGGAAAGATGATGATCAGACAGAGTTTTCATGAT 1530  
 DB 439 AAGATCAGCATAAACAGAACAGGAAAGATGATGATCAGACAGAGTTTTCATGAT 380  
 QY 1531 TCATCAAAACAGATCCATGAAGAAGAGACGCAAGAGAGAAATTCAGATGTTGAC 1590  
 DB 379 TCATCAAAACAGATCCATGAAGAAGAGACGCAAGAGAGAAATTCAGATGTTGAC 320  
 QY 1591 CAGCAGAAAGCTGCCAAGGTGTTGGCCAGCAGAGAGAACTTAATCCCTTAGAAT 1650  
 DB 319 CAGCAGAAAGCTGCCAAGGTGTTGGCCAGCAGAGAGAACTTAATCCCTTAGAAT 260  
 QY 1651 GACGATTCGGAAGAGAGCTGAGAGATGTCAGATTCATCGAGTTTCAAGAGAAAG 1710  
 DB 259 GACGATTCGGAAGAGAGCTGAGAGATGTCAGATTCATCGAGTTTCAAGAGAAAG 200  
 QY 1711 ATGAGAGAGTTTGTGAAGAGAGAGATGCTGATTAAGATCAAGAGAAAGATGAA 1770  
 DB 199 ATGAGAGAGTTTGTGAAGAGAGAGATGCTGATTAAGATCAAGAGAAAGATGAA 140

QY 1771 GACATGAGAGAGAGCATCAGAGAGATATT 1803  
 DB 139 GACATGAGAGAGAGCATCAGAGAGATATT 107

RESULT 4  
 LOCUS AU236368 647 bp mRNA linear EST 01-APR-2002  
 DEFINITION AU236368 RAFL14 Arabidopsis thaliana cDNA clone RAFL14-93-K05 5',  
 mRNA sequence.  
 ACCESSION AU236368  
 VERSION AU236368.1 GI:19875537  
 KEYWORDS EST.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 647)  
 Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Saitou, M., Nakajima, M.,  
 Oono, Y., Sakurai, T., Shimada, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.,  
 Arakawa, T., Shibata, K., Kaminari, P., Kawai, J., Itoh, M., Ichi, Y.,  
 and Shinohara, K.  
 Large scale analysis of Arabidopsis full-length cDNA (2002b)  
 Unpublished (2002)  
 Contact: Motoaki Seki  
 Plant Functional Genomics Research Group  
 RIKEN Genomic Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-4359  
 Fax: 81-298-36-9060  
 Email: msek@rcc.riken.go.jp  
 An Arabidopsis full-length cDNA library was constructed essentially  
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI  
 and XhoI was ligated to modified lambda FLIC-1 vector (Garnini et  
 al., submitted for publication) digested with BamHI and SalI. This  
 clone is in a modified pBluescript vector. Please visit our web  
 site (http://www.gsc.riken.go.jp/e/plant/index\_e.html) for further  
 details.

ORIGIN  
 Query Match 27.3%; Score 512; DB 9; Length 647;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-249;  
 Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTCTTAGGCTGCTCAATGCTTAAGAGAAAGAAAGAGTTCAGAGGTGTTAGGCT 60  
 DB 92 ATGAGTCTTAGGCTGCTCAATGCTTAAGAGAAAGAAAGTTCAGAGGTGTTAGGCT 151  
 QY 61 GAGGTGAACAGTGTGCTCAAGTTTGGAGGAGCAGACTGCTTCTCAAGATGAT 120  
 DB 152 GAGGTGAACAGTGTGCTCAAGTTTGGAGGAGCAGACTGCTTCTCAAGATGAT 211  
 QY 121 GGAGGAGAGTGGAGGTCAATTTCAAGAGAAACAGAACAAACAGAGAACTTCTGA 180  
 DB 212 GGAGGAGAGTGGAGGTCAATTTCAAGAGAAACAGAACAAACAGAGAACTTCTGA 271  
 QY 181 AAAAATTGGGTTTTCAGAAATTCGAATCTCTAGAGCTTGGGCTGTCAGAGCAAGGG 240  
 DB 272 AAAAATTGGGTTTTCAGAAATTCGAATCTCTAGAGCTTGGGCTGTCAGAGCAAGGG 331  
 QY 241 AGAGTAGCAAGTATCTGAGAGAGAAACAATGTATCCGAGAGAGTAACGGCAATGCT 300



Db 332 AGAGGTAGCAAGTATCTGGAGAGAAACATGATCCGAGAGGTAAAGGAAATGAT 391  
 QY 301 CGGGGCACTTCAAGCTTACATCTGTGGGGAGAGGCTTACAGCAAAATGATATAC 360  
 Db 392 CGGGGCACTTCAAGCTTACATCTGTGGGGAGAGGCTTACAGCAAAATGATATAC 451  
 QY 361 AACTTGTGGCAACCCCACTGATCTCGCCCTCTTTGAAAGAGATGAAATTTGGAG 420  
 Db 452 AACTTGTGGCAACCCCACTGATCTCGCCCTCTTTGAAAGAGATGAAATTTGGAG 511  
 QY 421 GCAAGAGAGGCTTCTGTGACGACACAGCTGAGGAGATTTCTGACGTGGAGATAT 480  
 Db 512 GCAAGAGAGGCTTCTGTGACGACACAGCTGAGGAGATTTCTGACGTGGAGATAT 571  
 QY 481 GTGATATATGCTTCTGAGAAAGAAATGATTC 512  
 Db 572 GTGATATATGCTTCTGAGAAAGAAATGATTC 603

RESULT 5 613 bp DNA linear GSS 28-JAN-2000  
 LOCUS AQ964580/c Arabidopsis thaliana genomic clone LERGX20, genomic  
 DEFINITION LERGX20TF LERG Arabidopsis thaliana genomic clone LERGX20, genomic  
 survey sequence.  
 ACCESSION AQ964580  
 VERSION AQ964580.1 GI:6792281  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 613)  
 Buell, C.R., Jin, X., Pai, G., Barnstead, M., Bowman, C., Utechtach, T.,  
 Feldblum, T., Liang, F., Creasy, T. and Frazer, C.M.  
 Genomic survey sequencing of Landsberg erecta ecotype of  
 Arabidopsis thaliana and identification of sequence-based  
 polymorphisms  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Xiaoyang Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: at@tigr.org  
 For additional information, see <http://www.tigr.org/cdb/at/cdb.html>  
 Seq primer: TF  
 Class: shotgun.  
 FEATURES  
 source location/Qualifiers

1..613  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="LANDSBERG ERECTA"  
 /db\_xref="taxon:3702"  
 /clone="LERGX20"  
 /note="Organ: Leaf; Vector: pUC19UK. Total genomic DNA was  
 sheared to 0.4-0.7 kbp before ligation."

ORIGIN  
 Query Match 25.2%; Score 474; DB 28; Length 613;  
 Best Local Similarity 99.8%; Pred. No. 4.5e-220; Indels 0; Gaps 0;  
 Matches 524; Conservative 1;

QY 1 ATGAGTTAGAGGCTGCTCAATGCTTAAGAAAGAAAGCTTACAGGCTGTATAGGCTT 60  
 Db 525 ATGAGTTAGAGGCTGCTCAATGCTTAAGAAAGAAAGCTTACAGGCTGTATAGGCTT 466  
 QY 61 GAGGTTGAACAGTTGCTTCAAGTTGGCAGGAGCGAGACTGGCTTCTTCAAGATGAT 120  
 Db 465 GAGGTTGAACAGTTGCTTCAAGTTGGCAGGAGCGAGACTGGCTTCTTCAAGATGAT 406

QY 121 GGAGAGAGGTGGAGGTCTTTCCAGAGAAAGAAAGCAACAGAGAAACACTTCTGA 180  
 Db 405 GGAGAGAGGTGGAGGTCTTTCCAGAGAAAGAAAGCAACAGAGAAACACTTCTGA 346  
 QY 181 AAAAAGTGGGTTTCCAGAAATTCGAATCTCTTGAAGCTTGGGGTGTCTGACGAAAGG 240  
 Db 345 AAAAAGTGGGTTTCCAGAAATTCGAATCTCTTGAAGCTTGGGGTGTCTGACGAAAGG 286  
 QY 241 AGAGGTAGCAACATATCTGTGGAGAGAGAAACAATGATCCGGAGAGGTTAACGCAATGCT 300  
 Db 285 AGAGGTAGCAACATATCTGTGGAGAGAGAAACAATGATCCGGAGAGGTTAACGCAATGCT 226  
 QY 301 CGGGGCACTTCAAGCTTACATCTGTGGAGAGAGAAACAATGATCCGGAGAGGTTAAC 166  
 Db 225 CGGGGCACTTCAAGCTTACATCTGTGGAGAGAGAAACAATGATCCGGAGAGGTTAAC 106  
 QY 361 AACTTGTGGCAACCCCACTGATCTCGCCCTCTTTGAAAGAGATGAAATTTGGAG 420  
 Db 165 AACTTGTGGCAACCCCACTGATCTCGCCCTCTTTGAAAGAGATGAAATTTGGAG 106  
 QY 421 GCAAGAGAGGTTCTGTGACGACACAGCTGAGAGATTTCTGACGTGAGAGATAT 480  
 Db 105 GCAAGAGAGGTTCTGTGACGACACAGCTGAGAGATTTCTGACGTGAGAGATAT 46  
 QY 481 GTGATATATGCTTCTGAGAAAGAAATGATTCGATCTTTGGAT 525  
 Db 45 GTGATATATGCTTCTGAGAAAGAAATGATTCGATCTTTGGAT 1

RESULT 6 380 bp mRNA linear EST 01-SEP-2000  
 LOCUS AV525508 Arabidopsis thaliana aboveground organs two to six-week  
 DEFINITION old Arabidopsis thaliana cDNA clone APD25402R 5', mRNA sequence.  
 ACCESSION AV525508  
 VERSION AV525508.1 GI:8685036  
 KEYWORDS EST.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 380)  
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries  
 JOURNAL DNA Res. 7 (3), 175-180 (2000)  
 MEDLINE 20363093  
 PUBMED 10907847

COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>  
 FEATURES  
 source location/Qualifiers

1..380  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="APD25402R"  
 /tissue\_type="aboveground organs"  
 /dev\_stage="two to six-week old"  
 /clone\_1lb="Arabidopsis thaliana aboveground organs two to  
 six-week old"  
 /note="Vector: pBluescriptII SK-, Site\_1: EcoRI; Site\_2:  
 XhoI"

ORIGIN  
 Query Match 17.5%; Score 329; DB 9; Length 380;  
 Best Local Similarity 99.7%; Pred. No. 4e-156;  
 Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 276 ATCCGGAGAGGTAAACGGCAATGTCGGGCGATTCAGCTAATATCTGTCGGGACG 335  
 DB 1 ATCCGGAGAGGTAAACGGCAATGTCGGGCGATTCAGCTAATATCTGTCGGGACG 60  
 QY 336 AGCGTTGAGCAGAAAGATATATTAACAATTGTCGACCCCCACCTGATCTGCTCC 395  
 DB 61 AGCGTTGAGCAGAAAGATATATTAACAATTGTCGACCCCCACCTGATCTGCTCC 120  
 QY 396 TTGGAAGAGAGATGGAATTTGGCAGGCAAGAGGTTCTGTCAGCAGCTGTCGA 455  
 DB 121 TTGGAAGAGAGATGGAATTTGGCAGGCAAGAGGTTCTGTCAGCAGCTGTCGA 180  
 QY 456 GGAGTTTCTGACGTGAGAGATGATGATTAATGCTTCTGAGGAAGAGATATTCGA 515  
 DB 181 GGAGTTTCTGACGTGAGAGATGATGATTAATGCTTCTGAGGAAGAGATATTCGA 240  
 QY 516 TGCCTTGGATGATTCGTGATGACGACCTTGCAAGTATGATTAATGCTGAGTGTGACTCA 575  
 DB 241 TGCCTTGGATGATTCGTGATGACGACCTTGCAAGTATGATTAATGCTGAGTGTGACTCA 300  
 QY 576 AAGAGCCATGATTCAGAAAGCAAGATAGTGTTCAAAAAGTTCTTGGCAGCTTGA 635  
 DB 301 AAGAGCCATGATTCAGAAAGCAAGATAGTGTTCAAAAAGTTCTTGGCAGCTTGA 360  
 QY 636 TAGCTTGTGATGACGACGA 655  
 DB 361 TAGCTTGTGATGACGACGA 380

RESULT 7  
 CENS00P9H 457 bp DNA linear GSS 28-JUN-1999  
 LOCUS Arabidopsis thaliana genome survey sequence T7 end of BAC F8G21 of  
 DEFINITION IGF library from strain Columbia of Arabidopsis thaliana, genomic  
 survey sequence.  
 AL084227 GI:5285367  
 GSS.

ACCESSION Arabidopsis thaliana (thale cress)  
 VERSION Arabidopsis thaliana  
 KEYWORDS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 SOURCE 1 (bases 1 to 457)  
 ORGANISM Salanoubat, M., Choisme, N., Artiguenave, F., Brotier, P., Wincker, P.,  
 Samson, D., Saurin, W., Weissenbach, J., and Quetier, F.

REFERENCE JOURNAL 2 (bases 1 to 457)  
 AUTHORS Unpublished  
 TITLE Genoscope.  
 REFERENCE Direct Submission  
 JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

FEATURES  
 source 1..457  
 Location/Qualifiers

/organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_1ib="F8G21"  
 /clone\_1ib="IGF"  
 /note="end : T7"

ORIGIN  
 Query Match 15.7%; Score 295; DB 29; Length 457;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-139;  
 Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 962 AGTGGCTCGGCAAGGCAACAGAGCTGCTGAATCTTGCACAGATAGAGGCTTTA 1021  
 DB 55 AGTGGCTCGGCAAGGCAACAGAGCTGCTGAATCTTGCACAGATAGAGGCTTTA 114

QY 1022 GAGCAGCCATTCCTATGCTCCACAGGCGCATCTGGAGTGAAGTCTTGTATGTTGAGA 1081  
 DB 115 GAGCAGCCATTCCTATGCTCCACAGGCGCATCTGGAGTGAAGTCTTGTATGTTGAGA 174  
 QY 1082 GCAGTCCACCTGCTATTTTGGAGCCGAAACGCTCCACCGGAGTTAGCTGAGATGGGT 1141  
 DB 175 GCAGTCCACCTGCTATTTTGGAGCCGAAACGCTCCACCGGAGTTAGCTGAGATGGGT 234  
 QY 1142 TAGATGAATTTGCTGGGCTCAGAAAGCAGATATGTTTCTGGAAGGTTTCCCACTGT 1201  
 DB 235 TAGATGAATTTGCTGGGCTCAGAAAGCAGATATGTTTCTGGAAGGTTTCCCACTGT 294  
 QY 1202 ATGCTTCTTTCACAGCAAGAGATCTGACATATTCATCAACTCTCAAG 1256  
 DB 295 ATGCTTCTTTCACAGCAAGAGATCTGACATATTCATCAACTCTCAAG 349

RESULT 8  
 A0959659 523 bp DNA linear GSS 28-JAN-2000  
 LOCUS LEREM21TR LERE Arabidopsis thaliana genomic clone LEREM21, genomic  
 DEFINITION survey sequence.  
 A0959659  
 VERSION A0959659.1 GI:6787360  
 KEYWORDS GSS.

ACCESSION Arabidopsis thaliana (thale cress)  
 VERSION Arabidopsis thaliana  
 KEYWORDS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 SOURCE 1 (bases 1 to 523)  
 ORGANISM Buell, C.R., Lin, X., Pai, G., Barnes, M., Bowman, C., Utebach, T.,  
 Feldlyum, T., Liang, F., Creasy, T., and Fraser, C.M.

REFERENCE JOURNAL 1 (bases 1 to 523)  
 AUTHORS Genomic survey sequencing of Landsberg erecta ecotype of  
 Arabidopsis thaliana and identification of sequence-based  
 polymorphisms  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: atc@igr.org  
 For additional information, see <http://www.tigr.org/cdb/at.html>  
 Seq primer: TR  
 Class: shotgun.

FEATURES  
 source 1..523  
 Location/Qualifiers

/organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="LANDSBERG ERRECTA"  
 /db\_xref="taxon:3702"  
 /clone\_1ib="LEREM21"  
 /clone\_1ib="LERE"  
 /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was  
 sheared to 0.6-0.8 Kbp before ligation."

ORIGIN  
 Query Match 14.5%; Score 273; DB 28; Length 523;  
 Best Local Similarity 99.3%; Pred. No. 1.8e-127;  
 Matches 423; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 167 GAAACACTTCTGAAAAAATCTGGGTTTCTCAGAAATTCGAATCTCTTAAGCTTTGGGCTG 226  
 DB 25 GAAACACTTCTGAAAAAATCTGGGTTTCTCAGAAATTCGAATCTCTTAAGCTTTGGGCTG 84

QY 227 GTCAGCAGCAAGAGAGATGACCAAGTATCTGGAGAGAGAAACAATGTATCCGAGAG 286  
 DB 85 GTCAGCAGCAAGAGAGATGACCAAGTATCTGGAGAGAGAGAAACAATGTATCCGAGAG 144

QY 287 GTAACGGCAATGCTGCGGCGCATTCAGGTAATATCTGCTGCGGAGCAGCGTTGAGCA 346  
 DB 145 GTAACGGCAATGCTGCGGCGCATTCAGGTAATATCTGCTGCGGAGCAGCGTTGAGCA 204

QY	347	GAAGATATGATAACAATTCTTGTGGCACCACCACTGTATCTGGGCTCCTTTGGAAAGAG	408
Db	205	GAAGATATGATAACAATTCTTGTGGACCCCACTGTATCTGGGCTCCTTTGGAAAGAG	264
QY	407	GATGGAATTTGGCAGGCAAGAGAGGTTCTGCTCAGCACACAGCTGTGAGAGTTTCTGT	466
Db	265	GATGGAATTTGGCAGGCAAGAGAGGTTCTGCTCAGCACACAGATGTGAGAGTTTCTGT	324
QY	467	ACGCGGAGGATGATGTGTGATATGCTTTGAGAGAAAGATGATTTCCGATCTTTGGATG	526
Db	325	ACGCGGAGGATGATGTGTGATATGCTTTGAGAGAAAGATGATTTCCGATCTTTGGATG	384
QY	527	ATTCTGATGACGACCTTGCAAGTGATGATTTATGACTCGGATGTAGTCAAAAGGCCATG	586
Db	385	ATTCTGATGACGACCTTGCAAGTGATGATTTATGACTCGGATGTAGTCAAAAGGCCATG	444
QY	587	GATCAC 592	
Db	445	GATCAC 450	

RESULT 9					
AU227277/c					
LOCUS	AU227277	421 bp	mRNA	linear	EST 23-APR-2007
DEFINITION	AU227277 RAFL14	Arabidopsis	thaliana	CDNA	clone RAFL14-93-K05 3',
	mRNA sequence.				

ACCESSION	AU227277
VERSION	AU227277.1
	GI:19741924

KEYWORDS	EST.
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana

REFERENCE  
AUTHORS

1 (Baesens et al. 421),  
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M.,  
Kawai, T., Kato, Y., Yamada, T., Ohguchi, S., Ogasawara, P.  
Kawai, T.

TITLE  
JOURNAL  
large scale analysis of Arabidopsis full-length cDNA  
unpublished (2002)

COMMENT

Contact: MOTOAKI SAKI  
Plant Functional Genomics Research Group  
RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel.: 81-298-36-4359  
Fax: 81-298-36-0060

fax: 01-230-38-2001  
Email: mseki@cc.riken.go.jp  
An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda FL-C1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site ([http://www.gsc.riken.go.jp/e/plant/index\\_e.html](http://www.gsc.riken.go.jp/e/plant/index_e.html)) for further details.

```

FEATURES
source
1. 421
location/Qualifiers
1. /organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAF1.4-93-K05"
/tissue_type="root"
/lab_host="DH10B"
/clone_lib="RAFL14"
/note="Site_1: BamHI; Site_2: SalI"

```

Query Match 13.0%; Score 244; DB 9; Length 421;  
Best Local Similarity 100.0%; Pred. No. 1e-112;  
Matches 244; Conservative 0; Mismatches 0; Gaps 0;  
Indels 0

[illegible]

RESULT 10	
A0010650	
LOCUS	A0010650
DEFINITION	290 bp DNA linear GSS 29-MAY-1998
FEATURES	F27C8T8TC IGF Arabidopsis thaliana genomic clone F27C8, genomic survey sequence.

ACCESSION	AQ010650
VERSION	AQ010650.1
	GI:3165927

KEYWORDS	GSS.
SOURCE	Arabidopsis thaliana (thale cress)
ORGANISM	Arabidopsis thaliana

REFERENCE

Bounejev S. D., Shih F. J., Wible, C., Golden, K., Shateman, S., Choi, P., 1 (bases 1 to 290)

Eukaryota; Viridiplantae; Streptophyta; Charophyceae; Equisetophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eunotids II; Brassicales; Brassicaceae; Arabidopsids.

AUTHORS	Yun,K., Akinteye,B., Shen,K., Gonssekaram,S., Millscher,J., Adams,M.D. and Venter,J.C.
TITLE	A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing. Update 4
JOURNAL	Unpublished (1998)
COMMENT	Other_GS8s: F27C8TFC

Contact: Steve Rounsley  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: rounsley@tigr.org  
Seq primer: M13 Reverse  
Class: BAC ends  
High quality sequence stop: 290.  
Location/Qualifiers

```

FEATURES
  source
    1..230
    location=Genomic
    /organism="Arabidopsis thaliana"
    /mol_type="genomic DNA"
    /strain="Columbia"
    /db_xref="taxon:3702"
    /clone="F27C8"
    /sex="thermaproditea"
    /clone_1db="rigp"
    /note="Vector: Belobact1, Site_1: EcoRI, Site_2: EcoRI
    Produced by Thomas Altmann"

```

Query Match	12.1%	Score 228	DB 28;	Length 290;
Best Local Similarity	100.0%	Pred. No. 1.3e-104;		
Matches 228; Conservative	0;	Mismatches	0;	Gaps 0.

[illegible]

Db 123 GACACGCCCATCTCTATGCTCCACAGGCCATCGTGGATGAGTCTGATGTTGAGA 182

Qy 1082 GCAGTGCACCTGCTATTTGAGGCGGAGCCCTCCACGGGAGTTAGCTGAGATGGGGT 1141

Db 183 GCAGTGCACCTGCTATTTGAGGCGGAGCCCTCCACGGGAGTTAGCTGAGATGGGGT 242

Qy 1142 TAGATGAATTTGCTGGGGTCCAGAGCGCAGTATGTTTCTGAGAGTG 1189

Db 243 TAGATGAATTTGCTGGGGTCCAGAGCGCAGTATGTTTCTGAGAGTG 290

RESULT 11  
AI999551 548 bp mRNA linear EST 08-SEP-1999  
LOCUS 701556368 A. thaliana, Columbia Col-0, rosette-3 Arabidopsis  
DEFINITION thaliana cDNA clone 701556368, mRNA sequence.  
ACCESSION AI999551  
VERSION AI999551.1 GI:5846456  
KEYWORDS EST.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 548)  
AUTHORS Chen, J., Montiyama, M., Chan, E., Mooney, M., Carron, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Georgene, G., Burns, D., Griffin, J., Mounauyrou, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobriga, A., Murry, L., Turner, C., Kikorian, S., Elder, L. and Hanson, D.  
TITLE Arabidopsis thaliana Gene Expression Microarray  
JOURNAL Unpublished (1999)  
COMMENT Contact: David Smolter, Ph.D.  
Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc.  
4633 World Parkway Circle, St. Louis, MO 63134, USA  
Tel: 877-577-2733  
Fax: 314-427-3324  
Email: service@genomesystems.com.  
LOCATION/Qualifiers  
1. 548  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/cultivar="Columbia Col-0"  
/db\_xref="taxon:3702"  
/clone="701556368"  
/tissue\_type="rosette"  
/dev\_stage="4 - 7 weeks"  
/clone\_lib="A. thaliana, Columbia Col-0, rosette-3"  
/note="Vector: pSPORT; Site 1: NotI; Site 2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was bluntended, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

ORIGIN  
Query Match 9.0%; Score 169; DB 9; Length 548;  
Best Local Similarity 100.0%; Pred. No. 2.2e-74;  
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1710 GATGAGGAGCTTTGTGGAAGAGGAGATGCTGATAAAGATCAAGAGAAAGATGGA 1769

Db 345 GATGAGGAGCTTTGTGGAAGAGGAGATGCTGATAAAGATCAAGAGAAAGATGGA 286

Qy 1770 AGACATGAAGAAGAGCATCAGAGAGATATTGATCTGAGAAAGATTTGATGAGGC 1829

Db 285 AGACATGAAGAAGAGCATCAGAGAGATATTGATCTGAGAAAGATTTGATGAGGC 226

Qy 1830 TTGGAAGAGCTTCATGATACAGATGCGCTTCAATGAATGATGATGA 1878

Db 225 TTGGAAGAGCTTCATGATACAGATGCGCTTCAATGAATGATGATGA 177

RESULT 12  
BH169457 197 bp DNA linear GSS 03-OCT-2001  
LOCUS SALK\_001394 Arabidopsis thaliana TDNA insertion lines Arabidopsis  
DEFINITION thaliana genomic clone SALK\_001394, genomic survey sequence.  
ACCESSION BH169457  
VERSION BH169457.1 GI:15904832  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 197)  
AUTHORS Alonso, J.M., Leisbe, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shim, P., Zimmerman, J., and Ecker, D.R.  
TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome  
JOURNAL Unpublished (2001)  
COMMENT Contact: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGNAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: ecker@salk.edu  
This is single pass sequence recovered from the left border of TDNA.  
FEATURES  
source Location/Qualifiers  
1. 197  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="SALK\_001394"  
/clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

ORIGIN  
Query Match 8.1%; Score 153; DB 28; Length 197;  
Best Local Similarity 100.0%; Pred. No. 2.3e-66;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 50 GTTATAGGCTTGAAGTGTGTTCAAGTGTGTCAGAGGAGAGACTGGCTTCTT 109

Db 1 GTTATAGGCTTGAAGTGTGTTCAAGTGTGTCAGAGGAGAGACTGGCTTCTT 60

Qy 110 CACAAGATGATGAGAGAGAGTGGAGGTGATTTCCAGAAGAAACAAGAACCAAGGAA 169

Db 61 CACAAGATGATGAGAGAGAGTGGAGGTGATTTCCAGAAGAAACAAGAACCAAGGAA 120

Qy 170 ACACTTCTGAAAAAATTGGGTTTCTCAGAAAT 202

Db 121 ACACTTCTGAAAAAATTGGGTTTCTCAGAAAT 153

RESULT 13  
BH618393/c

LOCUS BH618393 443 bp DNA linear GSS 30-JAN-2002  
 DEFINITION SALK\_030005 Arabidopsis thaliana TDNA insertion lines Arabidopsis  
 thaliana genomic clone SALK\_030005, genomic survey sequence.  
 ACCESSION BH618393  
 VERSION BH618393.1 GI:18428488  
 KEYWORDS GSS.  
 ORGANISM Arabidopsis thaliana (thale cress)  
 SOURCE Arabidopsis thaliana  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 443)  
 Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,  
 Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,  
 Shinn, P., Zimmerman, J., and Ecker, J.R.  
 A sequence-indexed library of insertion mutations in the  
 Arabidopsis Genome  
 Unpublished (2001)  
 CONTACT: Joseph R. Ecker  
 Salk Institute Genomic Analysis Laboratory (SIGAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6379  
 Email: eckersalk.salk.edu  
 This is single pass sequence recovered from the left border of  
 TDNA. This sequence lies within an annotated intron of At5g23570.  
 Class: TDNA tagged.  
 Location/Qualifiers  
 1..443  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
 /note="PCR was performed on Arabidopsis thaliana lines  
 each of which contains one or more TDNA insertion  
 elements. The resultant fragment for each line was  
 directly sequenced to determine the genomic sequence at  
 the site of insertion. Details of the protocols used can  
 be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"  
 ORIGIN  
 Query Match 7.9%; Score 148; DB 28; Length 443;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-63;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 962 AGTGGCTGGGATGGGACACCAAGAGTGTGATATCTGACCAAGTATGAGGCTCTTA 1021  
 |||||  
 Db 407 AGTGGCTGGGATGGGACACCAAGAGTGTGATATCTGACCAAGTATGAGGCTCTTA 348  
 |||||  
 Oy 1022 GAGCAACGCGATTCCTATGTCACACAGGCGCATGCGGATGAGTGTTCGATGTTTGACA 1081  
 |||||  
 Db 347 GAGCAACGCGATTCCTATGTCACACAGGCGCATGCGGATGAGTGTTCGATGTTTGACA 288  
 |||||  
 Oy 1082 GCAGTGCCACATGCTATTGAGGCCCA 1109  
 |||||  
 Db 287 GCAGTGCCACATGCTATTGAGGCCCA 260  
 |||||  
 RESULT 14  
 BH169448 482 bp DNA linear GSS 03-OCT-2001  
 LOCUS BH169448  
 DEFINITION SALK\_001377 Arabidopsis thaliana TDNA insertion lines Arabidopsis  
 thaliana genomic clone SALK\_001377, genomic survey sequence.  
 ACCESSION BH169448  
 VERSION BH169448.1 GI:15904823  
 KEYWORDS GSS.  
 ORGANISM Arabidopsis thaliana (thale cress)  
 SOURCE Arabidopsis thaliana  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 482)  
 Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R.,  
 Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L.,  
 Shinn, P., Zimmerman, J., and Ecker, J.R.  
 A sequence-indexed library of insertion mutations in the  
 Arabidopsis Genome  
 Unpublished (2001)  
 CONTACT: Joseph R. Ecker  
 Salk Institute Genomic Analysis Laboratory (SIGAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6379  
 Email: eckersalk.salk.edu  
 This is single pass sequence recovered from the left border of  
 TDNA.  
 Class: TDNA tagged.  
 Location/Qualifiers  
 1..482  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
 /note="PCR was performed on Arabidopsis thaliana lines  
 each of which contains one or more TDNA insertion  
 elements. The resultant fragment for each line was  
 directly sequenced to determine the genomic sequence at  
 the site of insertion. Details of the protocols used can  
 be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"  
 ORIGIN  
 Query Match 3.0%; Score 57; DB 28; Length 482;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-17;  
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 146 AGAGAGACAGACCAACCAAGCACTCTGGAACCACTGGGCTTCTCAGATT 202  
 |||||  
 Db 111 AGAGAGACAGACCAACCAAGCACTCTGGAACCACTGGGCTTCTCAGATT 167  
 |||||  
 RESULT 15  
 AO964581 232 bp DNA linear GSS 28-JAN-2000  
 LOCUS AO964581  
 DEFINITION LERGX20TR LERG Arabidopsis thaliana genomic clone LERGX20, genomic  
 survey sequence.  
 ACCESSION AO964581  
 VERSION AO964581.1 GI:6792282  
 KEYWORDS GSS.  
 ORGANISM Arabidopsis thaliana (thale cress)  
 SOURCE Arabidopsis thaliana  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE 1 (bases 1 to 232)  
 Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Utebach, T.,  
 Feldblum, T., Liang, F., Creasy, T., and Fraser, C.M.  
 Genomic survey sequencing of Landsberg erecta ecotype of  
 Arabidopsis thaliana and identification of sequence-based  
 polymorphisms  
 Unpublished (2000)  
 CONTACT: Xiaoying Lin  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: atel@igf.org  
 For additional information, see <http://www.tigr.org/cdb/ac/at.html>  
 Seq primer: TR  
 Class: shotgun.  
 Location/Qualifiers  
 FEATURES  
 JOURNAL  
 COMMENT  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: atel@igf.org  
 For additional information, see <http://www.tigr.org/cdb/ac/at.html>  
 Seq primer: TR  
 Class: shotgun.  
 Location/Qualifiers

source

```
1. .232
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="LANDSBERG_ERECTA"
/db_xref="taxon:3702"
/clone_lib="LBRGX20"
/notes="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
sheared to 0.4-0.7 kbp before ligation."
```

ORIGIN

```
Query Match          2.5%; Score 47; DB 28; Length 232;
Best Local Similarity 100.0%; Pred. No. 3.2e-12;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      6 TTCTAGGGCTGTCATGCTCTAAGGAAAGACGTTCAAGGTGTT 52
        |||||||
Db      134 TTCTAGGGCTGTCATGCTCTAAGGAAAGACGTTCAAGGTGTT 180
```

Search completed: April 9, 2004, 08:16:07  
Job time : 3161.03 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2004, 20:17:10 ; Search time 92.5698 Seconds  
(without alignments)  
11258.512 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878  
Sequence: 1 atgagcttcaggcgctgctcc.....ttccaatgaagatgatga 1878

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	1.0	481	US-09-615-192A-109	Sequence 109, App
C 2	19	1.0	481	US-09-169-789-109	Sequence 109, App
C 3	19	1.0	495	US-08-975-316-27	Sequence 27, Appl
C 4	19	1.0	495	US-09-615-192A-27	Sequence 27, Appl
C 5	19	1.0	495	US-09-169-789-27	Sequence 27, Appl
C 6	19	1.0	2000	US-09-041-075A-4	Sequence 4, Appl
C 7	19	1.0	2000	US-09-041-075A-6	Sequence 6, Appl
C 8	19	1.0	8257	US-09-595-684B-30	Sequence 30, Appl
C 9	19	1.0	8503	US-09-620-312D-130	Sequence 130, Appl
C 10	19	1.0	1664976	US-08-916-421B-1	Sequence 1, Appl
C 11	18	1.0	462	US-09-489-039A-5315	Sequence 5315, Ap
C 12	18	1.0	1254	US-09-322-478-25	Sequence 25, Appl
C 13	18	1.0	1829	US-07-668-352C-15	Sequence 15, Appl
C 14	18	1.0	1829	US-08-474-379C-15	Sequence 15, Appl
C 15	18	1.0	1829	US-09-146-249A-15	Sequence 15, Appl
C 16	18	1.0	1829	US-08-206-188B-15	Sequence 15, Appl
C 17	18	1.0	1829	PCT-US91-02714-15	Sequence 15, Appl
C 18	18	1.0	1952	US-09-533-029-89	Sequence 89, Appl
C 19	18	1.0	2001	US-09-489-039A-5217	Sequence 5217, Ap
C 20	18	1.0	3031	US-08-785-241-2	Sequence 15, Appl
C 21	18	1.0	7478	US-10-104-966-15	Sequence 15, Appl
C 22	18	1.0	8906	US-08-826-267-1	Sequence 1, Appl
C 23	18	1.0	35524	US-08-923-137-1	Sequence 1, Appl
C 24	18	1.0	42571	US-09-810-347-3	Sequence 3, Appl
C 25	18	1.0	100848	US-09-556-002-39	Sequence 39, Appl
C 26	17	0.9	20	US-09-851-896-62	Sequence 62, Appl
C 27	17	0.9	47	US-09-422-978-3322	Sequence 3322, Ap

28	17	0.9	247	4	US-09-411-977-5	Sequence 5, Appl
C 29	17	0.9	396	4	US-09-107-532A-767	Sequence 767, App
C 30	17	0.9	428	4	US-09-702-705-1214	Sequence 1214, Ap
C 31	17	0.9	428	4	US-09-736-457-1214	Sequence 1214, Ap
C 32	17	0.9	428	4	US-09-614-124B-1214	Sequence 1214, Ap
C 33	17	0.9	428	4	US-09-671-325-1214	Sequence 1214, Ap
C 34	17	0.9	477	4	US-09-621-976-10972	Sequence 10972, A
C 35	17	0.9	536	4	US-09-621-976-16220	Sequence 16220, A
C 36	17	0.9	561	1	US-08-565-386-3	Sequence 3, Appl
C 37	17	0.9	576	4	US-09-107-532C-1728	Sequence 1728, Ap
C 38	17	0.9	793	4	US-09-221-017B-940	Sequence 940, App
C 39	17	0.9	1250	3	US-09-018-584A-36	Sequence 36, Appl
C 40	17	0.9	1284	4	US-09-328-352-594	Sequence 594, App
C 41	17	0.9	1290	4	US-09-543-681A-1061	Sequence 1061, Ap
C 42	17	0.9	1323	4	US-09-489-039A-4961	Sequence 4961, Ap
C 43	17	0.9	1401	2	US-09-134-000C-2355	Sequence 2355, Ap
C 44	17	0.9	1428	4	US-08-960-022-707	Sequence 7, Appl
C 45	17	0.9	1496	4	US-09-016-434-1127	Sequence 1127, Ap

#### ALIGNMENTS

```
RESULT 1
US-09-615-192A-109/c
; Sequence 109, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Biokberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003CAU
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 109
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
; US-09-615-192A-109

Query Match          1.0%; Score 19; DB 4; Length 481;
Best Local Similarity 100.0%; Pred.No.12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      74 TGGTTCAAGGTTTGGCAGG 92
Db      470 TGGTTCAAGGTTTGGCAGG 452

RESULT 2
US-09-169-789-109/c
; Sequence 109, Application US/09169789
; Patent No. 6653528
; GENERAL INFORMATION:
; APPLICANT: Biokberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003C2
; CURRENT APPLICATION NUMBER: US/09/169,789
; CURRENT FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; EARLIER APPLICATION NUMBER: US 08/713,000
```



EARLIER FILING DATE: 1996-09-11  
; NUMBER OF SEQ ID NOS: 185  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 109  
; LENGTH: 481  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-09-169-789-109

Query Match 1.0%; Score 19; DB 4; Length 481;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTTCAAGGTTTGGCAGG 92  
DB 470 TGGTTCAAGGTTTGGCAGG 452

RESULT 3  
US-08-975-316-27/c  
; Sequence 27, Application US/08975316  
; Patent No. 5952486

GENERAL INFORMATION:  
; APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka  
; APPLICANT: and GRISON, Alastair W.  
; TITLE OF INVENTION: MATERIALS AND METHODS FOR  
; TITLE OF INVENTION: THE MODIFICATION OF PLANT LIGNIN CONTENT  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Ann W. Speckman  
; STREET: 2601 Elliott Avenue, Suite 4185  
; CITY: Seattle  
; STATE: WA USA  
; COUNTRY: USA  
; ZIP: 98121

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/975,316  
; FILING DATE:

CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/713,000  
; FILING DATE: September 11, 1996  
; ATTORNEY/AGENT INFORMATION:

NAME: SLENTH, Janet  
; REGISTRATION NUMBER: 37,007  
; REFERENCE/DOCKET NUMBER: 110000/1003C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-269-0565  
; TELEFAX: 206-269-0563

TELEX:  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 495 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-975-316-27

Query Match 1.0%; Score 19; DB 2; Length 495;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTTCAAGGTTTGGCAGG 92  
DB 484 TGGTTCAAGGTTTGGCAGG 466

RESULT 4

US-09-615-192A-27/c  
; Sequence 27, Application US/09615192A  
; Patent No. 6410718  
; GENERAL INFORMATION:

APPLICANT: Bloksberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; TITLE OF INVENTION: Modification of Plant Lignin Content  
; FILE REFERENCE: 11000.1003C4U  
; CURRENT APPLICATION NUMBER: US/09/615,192A  
; CURRENT FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 08/975,316  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: US 08/713,000  
; PRIOR FILING DATE: 1996-09-11  
; PRIOR APPLICATION NUMBER: US 09/169,789  
; PRIOR FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 405  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 27  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-09-615-192A-27

Query Match 1.0%; Score 19; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTTCAAGGTTTGGCAGG 92  
DB 484 TGGTTCAAGGTTTGGCAGG 466

RESULT 5  
US-09-169-789-27/c  
; Sequence 27, Application US/09169789  
; Patent No. 6653528

GENERAL INFORMATION:  
; APPLICANT: Bloksberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; TITLE OF INVENTION: Modification of Plant Lignin Content  
; FILE REFERENCE: 11000.1003C2  
; CURRENT APPLICATION NUMBER: US/09/169,789  
; CURRENT FILING DATE: 1998-10-09  
; EARLIER APPLICATION NUMBER: US 08/975,316  
; EARLIER FILING DATE: 1997-11-21  
; EARLIER APPLICATION NUMBER: US 08/713,000  
; EARLIER FILING DATE: 1996-09-11

NUMBER OF SEQ ID NOS: 185  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 27  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-09-169-789-27

Query Match 1.0%; Score 19; DB 4; Length 495;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTTCAAGGTTTGGCAGG 92  
DB 484 TGGTTCAAGGTTTGGCAGG 466

RESULT 6

US-09-041-075A-4  
; Sequence 4, Application US/09041075A  
; Patent No. H002022  
; GENERAL INFORMATION:  
; APPLICANT: Heidler, Steven A

APPLICANT: Radding, Jeffrey A  
TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI  
FILE REFERENCE: X-11242 Sequence List  
Patent No. H002022  
CURRENT APPLICATION NUMBER: US/09/041,075A  
CURRENT FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/043,591  
PRIOR FILING DATE: 1997-04-15  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 4  
LENGTH: 2000  
TYPE: DNA  
ORGANISM: Candida krusei  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (300)..(1739)  
US-09-041-075A-4

Query Match 1.0%; Score 19; DB 1; Length 2000;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 523 GATGATTCGATGACGACC 541  
DB 1560 GATGATTCGATGACGACC 1578

RESULT 7  
US-09-041-075A-6  
Sequence 6, Application US/09041075A  
Patent No. H002022  
GENERAL INFORMATION:  
APPLICANT: Heidler, Steven A  
TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI  
FILE REFERENCE: X-11242 Sequence List  
Patent No. H002022  
CURRENT APPLICATION NUMBER: US/09/041,075A  
CURRENT FILING DATE: 1998-03-10  
PRIOR APPLICATION NUMBER: 60/043,591  
PRIOR FILING DATE: 1997-04-15  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patentln Ver. 2.1  
SEQ ID NO 6  
LENGTH: 2000  
TYPE: RNA  
ORGANISM: Candida krusei  
US-09-041-075A-6

Query Match 1.0%; Score 19; DB 1; Length 2000;  
Best Local Similarity 73.7%; Pred. No. 12;  
Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 523 GATGATTCGATGACGACC 541  
DB 1560 GAUGAUUCGAGACGACC 1578

RESULT 8  
US-09-595-684B-30  
Sequence 30, Application US/09595684B  
Patent No. 6544766  
GENERAL INFORMATION:  
APPLICANT: Berard, Christophe  
APPLICANT: Ohashi, Cara  
APPLICANT: Sakowicz, Roman  
APPLICANT: Vaisberg, Eugeni  
APPLICANT: Wood, Kenneth  
APPLICANT: Yu, Ming  
TITLE OF INVENTION: Human kinesins and methods of producing  
TITLE OF INVENTION: and purifying human kinesins  
FILE REFERENCE: cytop036

CURRENT APPLICATION NUMBER: US/09/595,684B  
CURRENT FILING DATE: 2002-06-24  
PRIOR APPLICATION NUMBER: 09/295,612  
PRIOR FILING DATE: 2000-04-20  
NUMBER OF SEQ ID NOS: 105  
SOFTWARE: PatsSeq for Windows Version 4.0  
SEQ ID NO 30  
LENGTH: 8257  
TYPE: DNA  
ORGANISM: Human  
US-09-595-684B-30

Query Match 1.0%; Score 19; DB 4; Length 8257;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1802 TTGATCTGGAGAAAGATT 1820  
DB 2141 TTGATCTGGAGAAAGATT 2159

RESULT 9  
US-09-620-312D-130  
Sequence 130, Application US/09620312D  
Patent No. 6569662  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yunding  
APPLICANT: Wang, Dunru  
APPLICANT: Wang, Zhiwei  
APPLICANT: John Tillinghaast  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/09/620,312D  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1105  
SOFTWARE: pc\_fl\_genes Version 1.0  
SEQ ID NO 130  
LENGTH: 8503  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (91)..(8082)  
US-09-620-312D-130

Query Match 1.0%; Score 19; DB 4; Length 8503;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1802 TTGATCTGGAGAAAGATT 1820  
DB 2141 TTGATCTGGAGAAAGATT 2159

RESULT 10

US-08-916-421B-1  
Sequence 1, Application US/08916421B  
Patent No. 6503729  
GENERAL INFORMATION:  
APPLICANT: Bult et al.  
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus  
Patent No. 6503729  
TITLE OF INVENTION: jamaaschii  
FILE REFERENCE: PB275  
CURRENT APPLICATION NUMBER: US/08/916,421B  
CURRENT FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: US 60/024,428  
PRIOR FILING DATE: 1996-08-22  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1664976  
TYPE: DNA  
ORGANISM: Methanococcus jamaaschii  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (28222)..(28222)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (28257)..(28258)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84773)..(84773)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84808)..(84808)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (84812)..(84812)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98120)..(98120)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98159)..(98159)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98239)..(98239)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98266)..(98266)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (98343)..(98343)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (103998)..(103998)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (148948)..(148948)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (163385)..(163385)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (191989)..(191989)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (191995)..(191995)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (231980)..(231980)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234187)..(234187)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234220)..(234220)

OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (234814)..(234814)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309398)..(309398)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (309418)..(309418)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312837)..(312837)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (312993)..(312993)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (319226)..(319226)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (559167)..(559167)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (559241)..(559241)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (600992)..(600992)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (622708)..(622708)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (657081)..(657081)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (657203)..(657203)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (674435)..(674435)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (682442)..(682442)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (713652)..(713652)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (741684)..(741684)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (779455)..(779455)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (779676)..(779676)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (855539)..(855539)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1119881)..(1119881)  
OTHER INFORMATION: n equals a, t, c, or g  
NAME/KEY: misc feature  
LOCATION: (1130881)..(1130881)  
OTHER INFORMATION: n equals a, t, c, or g

```
/ NAME/KEY: misc_feature
/ LOCATION: (1310988)..(1310988)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1313224)..(1313224)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1349473)..(1349473)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1349491)..(1349491)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1470091)..(1470091)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1569020)..(1569020)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1602912)..(1602912)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1603734)..(1603734)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1637998)..(1637998)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc_feature
/ LOCATION: (1664854)..(1664855)
/ OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
```

```
Query Match 1.0%; Score 19; DB 4; Length 1664976;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1702 GAGAAAGAGTGGAGAGT 1720
Db 358045 GAGAAAGAGTGGAGAGT 358063
```

```
RESULT 11
US-09-489-039A-5315
/ Sequence 5315, Application US/09489039A
/ Patent No. 6610836
/ GENERAL INFORMATION:
/ APPLICANT: Gary Breton et. al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
/ FILE REFERENCE: 2709.2004001
/ CURRENT FILING DATE: 2000-01-27
/ PRIOR FILING DATE: 1999-01-29
/ NUMBER OF SEQ ID NOS: 14342
/ SEQ ID NO 5315
/ LENGTH: 462
/ TYPE: DNA
/ ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5315
```

```
Query Match 1.0%; Score 18; DB 4; Length 462;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 GCCAGCAGCAGCAGACA 1633
Db 216 GCCAGCAGCAGCAGACA 233
```

```
RESULT 12
US-09-322-478-25/c
/ Sequence 25, Application US/09322478
```

```
/ Patent No. 6331662
/ GENERAL INFORMATION:
/ APPLICANT: Wright, David A.
/ APPLICANT: Voytas, Daniel F.
/ TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
/ FILE REFERENCE: P-1065 ISURF Plant Retroelement
/ CURRENT APPLICATION NUMBER: US/09/322,478
/ EARLIER FILING DATE: 1999-05-28
/ EARLIER FILING DATE: 1998-05-29
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: Patentln Ver. 2.0
/ SEQ ID NO 25
/ LENGTH: 1254
/ TYPE: DNA
/ ORGANISM: Pisum sativum
US-09-322-478-25
```

```
Query Match 1.0%; Score 18; DB 4; Length 1254;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1335 TCAGCAGCTGACTACTT 1352
Db 890 TCAGCAGCTGACTACTT 873
```

```
RESULT 13
US-07-688-352C-15
/ Sequence 15, Application US/07688352C
/ Patent No. 5527896
/ GENERAL INFORMATION:
/ APPLICANT: Wiegler, Michael H.
/ APPLICANT: Colicelli, John J.
/ TITLE OF INVENTION: Cloning by Complementation and Related
/ NUMBER OF SEQUENCES: 57
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Marshall, O'Toole, Gerstein, Murray &
/ ADDRESSER: Bicknell.
/ STREET: Two First National Plaza, 20 South Clark
/ STREET: Street
/ CITY: Chicago
/ STATE: Illinois
/ COUNTRY: USA
/ ZIP: 60603
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: IBM PC compatible
/ SOFTWARE: Patentln Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/688,352C
/ FILING DATE: 19910419
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/511,715
/ FILING DATE: 20-APR-1990
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Borun, Michael F.
/ REGISTRATION NUMBER: 25447
/ REFERENCE/DOCKET INFORMATION:
/ TELECOMMUNICATION INFORMATION: 27805/30197
/ TELEPHONE: (312) 346-5750
/ TELEFAX: (312) 984-9740
/ TELEX: 25-3856
/ INFORMATION FOR SEQ ID NO: 15:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1829 base pairs
/ TYPE: NUCLEIC ACID
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
```

FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1421  
US-07-688-352C-15

Query Match 1.0%; Score 18; DB 1; Length 1829;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 ATTCCAGAGAACAG 156  
Db 985 ATTCCAGAGAACAG 1002

RESULT 14  
US-08-474-379C-15  
Sequence 15, Application US/08474379C  
Patent No. 5977305  
GENERAL INFORMATION:  
APPLICANT: Wigler, Michael H.  
APPLICANT: Colicelli, John J.  
TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive/6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,379C  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
INFORMATION FOR SEQ. ID NO: 15:  
APPLICATION DATA:  
FILING DATE: 20-APR-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/206,188  
FILING DATE: 01-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/688,352  
FILING DATE: 19-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 27866/32771  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
INFORMATION FOR SEQ. ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1829 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1421  
US-08-474-379C-15

Query Match 1.0%; Score 18; DB 2; Length 1829;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 ATTCCAGAGAACAG 156

Db 985 ATTCCAGAGAACAG 1002

RESULT 15  
US-09-146-249A-15  
Sequence 15, Application US/09146249A  
Patent No. 6069240  
GENERAL INFORMATION:  
APPLICANT: Wigler, Michael H.  
APPLICANT: Colicelli, John J.  
TITLE OF INVENTION: Cloning by Complementation and Related  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/146,249A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/511,715  
FILING DATE: 20-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W.  
REGISTRATION NUMBER: 36,107  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ. ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1829 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 30..1421  
US-09-146-249A-15

Query Match 1.0%; Score 18; DB 3; Length 1829;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 ATTCCAGAGAACAG 156  
Db 985 ATTCCAGAGAACAG 1002

Search completed: April 9, 2004, 08:20:46  
Job time : 96.5698 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2004, 05:52:11 ; Search time 529.543 Seconds  
(without alignments)  
13304.392 Million cell updates/sec

Title: US-10-030-829-2

Sequence: 1 atgagcttcaggcgtgtcc.....ttccaatgaagatgttga 1878

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 247585 seqs, 1875730760 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4951170

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	1.1	1377	12	US-10-424-599-91177 Sequence 91177, A
2	20	1.1	1589	12	US-10-424-599-91176 Sequence 91176, A
3	19	1.0	172	12	US-10-085-783A-19246 Sequence 19246, A
4	19	1.0	172	12	US-10-242-535A-19246 Sequence 19246, A
5	19	1.0	397	15	US-10-424-599-57837 Sequence 57837, A
6	19	1.0	429	15	US-10-027-632-293476 Sequence 293476, A
7	19	1.0	481	14	US-10-174-693-109 Sequence 109, App
8	19	1.0	495	14	US-10-174-693-27 Sequence 27, App
9	19	1.0	501	9	US-09-833-790-167 Sequence 167, App
10	19	1.0	681	12	US-10-424-599-18549 Sequence 18549, A
11	19	1.0	696	15	US-10-027-632-23555 Sequence 23555, A
12	19	1.0	696	15	US-10-027-632-23556 Sequence 23556, A
13	19	1.0	824	15	US-10-027-632-11811 Sequence 11811, A
14	19	1.0	824	15	US-10-027-632-11812 Sequence 11812, A
15	19	1.0	1137	15	US-10-369-493-24736 Sequence 24736, A

16	19	1.0	1341	9	US-09-938-842A-2465 Sequence 2465, Ap
17	19	1.0	1341	11	US-09-938-842A-2465 Sequence 2465, Ap
18	19	1.0	1576	12	US-10-424-599-4367 Sequence 4367, Ap
19	19	1.0	2000	9	US-09-742-582-4 Sequence 4, Appl
20	19	1.0	2000	9	US-09-742-582-6 Sequence 6, Appl
21	19	1.0	2000	10	US-09-742-580-4 Sequence 4, Appl
22	19	1.0	2000	10	US-09-742-580-6 Sequence 6, Appl
23	19	1.0	2000	10	US-09-742-581-4 Sequence 4, Appl
24	19	1.0	2000	10	US-09-742-581-6 Sequence 6, Appl
25	19	1.0	2240	15	US-10-108-260A-1451 Sequence 1451, Ap
26	19	1.0	3579	14	US-10-176-847-9 Sequence 9, Appl
27	19	1.0	6930	10	US-09-770-107-1 Sequence 1, Appl
28	19	1.0	7992	10	US-09-893-519A-140 Sequence 140, App
29	19	1.0	8491	14	US-10-133-013-260 Sequence 260, App
30	19	1.0	8493	13	US-10-071-766-51 Sequence 51, Appl
31	19	1.0	8503	14	US-10-037-270-130 Sequence 130, App
32	19	1.0	8503	15	US-10-117-722-130 Sequence 130, App
33	19	1.0	47448	15	US-10-085-117-145 Sequence 145, App
34	19	1.0	1691139	14	US-10-067-514-1 Sequence 1, Appl
35	19	1.0	1691139	15	US-10-419-723-1 Sequence 1, Appl
36	19	1.0	2940917	15	US-10-027-632-174763 Sequence 174763, A
37	18	1.0	182	14	US-10-029-386-26108 Sequence 26108, A
38	18	1.0	189	9	US-09-664-761-32008 Sequence 32008, A
39	18	1.0	209	9	US-09-560-352-12756 Sequence 12756, A
40	18	1.0	256	12	US-10-424-599-116017 Sequence 116017, A
41	18	1.0	270	9	US-09-960-352-14374 Sequence 14374, A
42	18	1.0	304	12	US-10-424-599-112786 Sequence 112786, A
43	18	1.0	312	9	US-09-728-445-290 Sequence 290, App
44	18	1.0	315	9	US-09-960-352-12961 Sequence 12961, A
45	18	1.0	395	9	US-09-960-352-13954 Sequence 13954, Ap

#### ALIGNMENTS

RESULT 1  
US-10-424-599-91177 Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223) B  
CURRENT APPLICATION NUMBER: US/10/424, 599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ. ID NOS: 285684  
SEQ. ID NO 91177  
LENGTH: 1377  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53343C.1  
US-10-424-599-91177

Query Match 1.1%; Score 20; DB 12; Length 1377;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 673 CAGTGCATTGTCACCTTG 692  
Db 359 CAGTGCATTGTCACCTTG 378  
RESULT 2  
US-10-424-599-91176  
Sequence 91176, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J

```

; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91176
; LENGTH: 1589
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1
; US-10-424-599-91176

Query Match          1.1%; Score 20; DB 12; Length 1589;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      673 CAGTGCATTGTCAGCTTG 692
DB      359 CAGTGCATTGTCAGCTTG 378

RESULT 3
US-10-085-783A-19246
; Sequence 19246, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: Liwew, C.C.
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19246
; LENGTH: 172
; TYPE: DNA
; ORGANISM: Human
; US-10-085-783A-19246

Query Match          1.0%; Score 19; DB 12; Length 172;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1767 GGAAGACATGAAGAAGAGG 1785
DB      86 GGAAGACATGAAGAAGAGG 104

RESULT 4
US-10-242-535A-19246
; Sequence 19246, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liwew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
```

```

; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19246
; LENGTH: 172
; TYPE: DNA
; ORGANISM: Human
; US-10-242-535A-19246

Query Match          1.0%; Score 19; DB 15; Length 172;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1767 GGAAGACATGAAGAAGAGG 1785
DB      86 GGAAGACATGAAGAAGAGG 104

RESULT 5
US-10-424-599-57837
; Sequence 57837, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 57837
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_23239C.1
; US-10-424-599-57837

Query Match          1.0%; Score 19; DB 12; Length 397;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1700 AAGAGAAAGATGAGAGA 1718
DB      262 AAGAGAAAGATGAGAGA 280

RESULT 6
US-10-027-632-293476
; Sequence 293476, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
```

; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 293476  
; LENGTH: 429  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-293476

Query Match 1.0%; Score 19; DB 15; Length 429;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 488 ATGCTTCTGAGGAAGAA 506  
DB 76 ATGCTTCTGAGGAAGAA 94

RESULT 7  
US-10-174-693-109/c  
; Sequence 109, Application US/10174693  
; Publication No. US20030131373A1  
; GENERAL INFORMATION:  
; APPLICANT: Blockberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; FILE REFERENCE: 11000.1003C5  
; CURRENT APPLICATION NUMBER: US/10/174,693  
; CURRENT FILING DATE: 2002-06-18  
; PRIOR APPLICATION NUMBER: US 08/975,316  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: US 08/713,000  
; PRIOR FILING DATE: 1996-09-11  
; PRIOR APPLICATION NUMBER: US 09/169,789  
; PRIOR FILING DATE: 1998-10-09  
; PRIOR APPLICATION NUMBER: US 09/615,192  
; PRIOR FILING DATE: 2000-07-12  
; NUMBER OF SEQ ID NOS: 407  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO: 109  
; LENGTH: 481  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-10-174-693-109

Query Match 1.0%; Score 19; DB 14; Length 481;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 TGGTCAAGGTTGGCAG 92  
DB 470 TGGTCAAGGTTGGCAG 452

RESULT 8  
US-10-174-693-27/c  
; Sequence 27, Application US/10174693  
; Publication No. US20030131373A1  
; GENERAL INFORMATION:  
; APPLICANT: Blockberg, Leonard N.  
; APPLICANT: Havukkala, Ilkka  
; TITLE OF INVENTION: Materials and Methods for the  
; FILE REFERENCE: 11000.1003C5  
; CURRENT APPLICATION NUMBER: US/10/174,693  
; CURRENT FILING DATE: 2002-06-18

; PRIOR APPLICATION NUMBER: US 08/975,316  
; PRIOR FILING DATE: 1997-11-21  
; PRIOR APPLICATION NUMBER: US 08/713,000  
; PRIOR FILING DATE: 1996-09-11  
; PRIOR APPLICATION NUMBER: US 09/169,789  
; PRIOR FILING DATE: 1998-10-09  
; PRIOR APPLICATION NUMBER: US 09/615,192  
; PRIOR FILING DATE: 2000-07-12  
; NUMBER OF SEQ ID NOS: 407  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO: 27  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-10-174-693-27

Query Match 1.0%; Score 19; DB 14; Length 495;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 TGGTCAAGGTTGGCAG 92  
DB 484 TGGTCAAGGTTGGCAG 466

RESULT 9  
US-09-833-790-167  
; Sequence 167, Application US/09833790  
; Patent No. US20020068288A1  
; GENERAL INFORMATION:  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Wang, Tonglong  
; APPLICANT: Secrist, Heather  
; APPLICANT: Mohamath, Raedeh  
; APPLICANT: Indrias, Carol Y.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.512  
; CURRENT APPLICATION NUMBER: US/09/833,790  
; CURRENT FILING DATE: 2001-04-11  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 167  
; LENGTH: 501  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-833-790-167

Query Match 1.0%; Score 19; DB 9; Length 501;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1802 TTGATCTGGAAGAAGATT 1820  
DB 10 TTGATCTGGAAGAAGATT 28

RESULT 10  
US-10-424-599-18549/c  
; Sequence 18549, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684



SEQ ID NO 18549  
LENGTH: 681  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_116754C.1  
US-10-424-599-18549

Query Match  
Best Local Similarity 100.0%; Score 19; DB 12; Length 681;  
Pred. No. 48;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1745 TAAAGATCGAGAGAGAA 1763  
DB 371 TAAAGATCGAGAGAGAA 353

RESULT 11  
US-10-027-632-23555/c  
Sequence 23555, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
TITLE OF INVENTION: Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 23555  
LENGTH: 696  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-23555

Query Match  
Best Local Similarity 100.0%; Score 19; DB 15; Length 696;  
Pred. No. 48;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1122 GGAGTTAGCTGAGATGGGG 1140  
DB 82 GGAGTTAGCTGAGATGGGG 64

RESULT 12  
US-10-027-632-23556/c  
Sequence 23556, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
TITLE OF INVENTION: Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 23556  
LENGTH: 696  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-23556

Query Match  
Best Local Similarity 100.0%; Score 19; DB 15; Length 696;  
Pred. No. 48;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1122 GGAGTTAGCTGAGATGGGG 1140  
DB 82 GGAGTTAGCTGAGATGGGG 64

RESULT 13  
US-10-027-632-11811  
Sequence 11811, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
TITLE OF INVENTION: Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11811  
LENGTH: 824  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-11811

Query Match  
Best Local Similarity 100.0%; Score 19; DB 15; Length 824;  
Pred. No. 48;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1491 CAGGAGAGATGATGCA 1509  
DB 579 CAGGAGAGATGATGCA 597

RESULT 14  
US-10-027-632-11812  
Sequence 11812, Application US/10027632

Job time : 534.543 secs

```

; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11812
; LENGTH: 824
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-11812

```

```

Query Match
Best local Similarity 100.0%; Score 19; DB 15; Length 824;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1491 CAGGGAAGAGATGATGCA 1509
DB 579 CAGGGAAGAGATGATGCA 597

```

```

RESULT 15
US-10-369-493-24736
; Sequence 24736, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfang
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 24736
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
US-10-369-493-24736

```

```

Query Match
Best local Similarity 100.0%; Score 19; DB 15; Length 1137;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1702 GAGAAAGAGATGAGAGT 1720
DB 607 GAGAAAGAGATGAGAGT 625

```

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:35:31 ; Search time 59 Seconds  
(without alignments)  
2993.088 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 625  
Sequence: 1 MRRAGPMSEKXVQGGYRP.....EPDEALQOLMYKHLNEDD 625

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	625	100.0	625	4 AAB31798	Aab31798 Amino aci
2	9	1.4	227	4 AAG91534	Aag91534 C glutam
3	9	1.4	237	4 AAB79177	Aab79177 Coryneb
4	8	1.3	483	3 AAB18189	Aab18189 Plasmodin
5	8	1.3	916	6 ABU23489	Abu23489 Protein e
6	8	1.3	1553	7 ADC99064	Adc99064 Human KPP
7	8	1.3	1572	5 AAE19162	Aae19162 Human kin
8	7	1.1	15	5 ABG71491	Abg71491 Prostati
9	7	1.1	18	6 ABP82841	Abp82841 G protein
10	7	1.1	32	4 AAE12564	Aae12564 p3serj/CI
11	7	1.1	34	4 AAO04467	Aao04467 Human pol
12	7	1.1	36	3 AAB13241	Aab13241 Caenorhab
13	7	1.1	43	5 AAO18208	Aao18208 Murine ro
14	7	1.1	43	5 AAO18207	Aao18207 Murine ro
15	7	1.1	43	5 AAU43846	Aau43846 Propionib
16	7	1.1	51	6 ABM40365	Abm40365 Propionib
17	7	1.1	57	5 ABP42713	Abp42713 Human ova
18	7	1.1	69	4 AAG74546	Aag74546 Human col
19	7	1.1	78	5 ABP08249	Abp08249 Human ORF
20	7	1.1	89	4 AAM89334	Aam89334 Human imm
21	7	1.1	90	4 AAU41632	Aau41632 Propionib
22	7	1.1	90	6 ABM38151	Abm38151 Propionib
23	7	1.1	99	5 ABP02735	Abp02735 Human ORF
24	7	1.1	103	4 AAU67643	Aau67643 Propionib
25	7	1.1	103	6 ABM64162	Abm64162 Propionib

26	7	1.1	103	6 ABM64798	Abm64798 Propionib
27	7	1.1	104	4 AAU21876	Aau21876 Human car
28	7	1.1	104	7 ADE34584	Ade34584 Human car
29	7	1.1	108	5 ABG71490	Abg71490 Human pro
30	7	1.1	123	4 AAU66188	Aau66188 Propionib
31	7	1.1	123	6 ABM62707	Abm62707 Propionib
32	7	1.1	124	4 ABG16117	Abg16117 Novel hum
33	7	1.1	138	6 ABP71171	Abp71171 Human pro
34	7	1.1	139	6 ABP71174	Abp71174 Chicken p
35	7	1.1	139	6 ABP71172	Abp71172 Rat pro-N
36	7	1.1	139	6 ABP71173	Abp71173 Mouse pro
37	7	1.1	141	6 ABU23918	Abu23918 Protein e
38	7	1.1	148	4 ABG19350	Abg19350 Novel hum
39	7	1.1	148	8 ADE52649	Ade52649 Murine dn
40	7	1.1	174	4 ABG29097	Abg29097 Novel hum
41	7	1.1	175	5 AAU99286	Aau99286 Rotavirus
42	7	1.1	182	4 AAU56433	Aau56433 Propionib
43	7	1.1	182	6 ABM52952	Abm52952 Propionib
44	7	1.1	211	4 AAB76573	Aab76573 Coryneb
45	7	1.1	213	6 ABP56866	Abp56866 Mouse CIT

#### ALIGNMENTS

RESULT 1	
AA31798	
ID AAB31798 standard; protein; 625 AA.	
XX	
AC AAB31798;	
DT 15-MAY-2001 (first entry)	
XX	
DE Amino acid sequence of the Arabidopsis SGS3 polypeptide.	
XX	
KW SGS3 gene; post-transcriptional inactivation; RNA degradation;	
KM viral resistance; resistance; fatty acid content; protein content.	
XX	
OS Arabidopsis thaliana.	
XX	
PN W0200105951.A2.	
XX	
PD 25-JAN-2001.	
XX	
PF 13-JUL-2000; 2000MO-FR002052.	
XX	
PR 16-JUL-1999; 99FR-00009417.	
PR 26-JAN-2000; 2000FR-00001006.	
XX	
PA (AVET ) AVENTIS CROPS SCIENCE SA.	
XX	
PT (INRG ) INST NAT RECH AGRONOMIQUE.	
XX	
DR Beclin C, Elmeyan T, Vaucheret H;	
XX	
DR WPI; 2001-159529/16.	
XX	
PT New SGS3 gene from Arabidopsis thaliana, useful for increasing virus	
PT resistance in plants and, when inhibited, for increasing transgene	
PT expression.	
XX	
PS Claim 8; Fig 1; 36pp; French.	
XX	
CC The present sequence represents an Arabidopsis thaliana SGS3 polypeptide.	
CC The SGS3 gene is essential for post-transcriptional inactivation	
CC (degradation of RNA) and for resistance to viruses. Overexpression of	
CC SGS3 results in plants with increased resistance to viruses, while	
CC inactivation of SGS3 in transgenic plants (e.g. by expressing antisense	
CC RNA, by mutation or by homologous recombination) increases the level of	
CC the transgene product. This product may e.g. impart resistance (to	
CC herbicide, insects or pathogens), alter contents of essential fatty acids	
CC or proteins, or is pharmaceutically active, e.g. an immunoglobulin or	
CC interferon	

XX Sequence 625 AA; 100.0%; Score 625; DB 4; Length 625;  
 SQ Query Match Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 625; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSSRAGPMSEKKNVQGGYRPEVEQIVQGLATGRLASSODDGGEMVYISKKNKPKGNTSG 60  
 DB 1 MSSRAGPMSEKKNVQGGYRPEVEQIVQGLATGRLASSODDGGEMVYISKKNKPKGNTSG 60  
 QY 61 KTVWSQNSNPRAWGGQQQGRGSGNVSGRNGNGRGIGIOMNISGRGRLSRKYDN 120  
 DB 61 KTVWSQNSNPRAWGGQQQGRGSGNVSGRNGNGRGIGIOMNISGRGRLSRKYDN 120  
 QY 121 NPVAPPVSRPPLGGNNVQARGSAQHTAVQGFPEVEDVDVNASSEENDSDALDDSD 180  
 DB 121 NPVAPPVSRPPLGGNNVQARGSAQHTAVQGFPEVEDVDVNASSEENDSDALDDSD 180  
 QY 181 LASDYSQNSQSHSGRKNKPKFPGSLDLSLBOINEPOROMHCPACONGPGLIDW 240  
 DB 181 LASDYSQNSQSHSGRKNKPKFPGSLDLSLBOINEPOROMHCPACONGPGLIDW 240  
 QY 241 YNLHPLAHARTKGRVYKHLRELAETLEKDLQWRGASVTPGSEIYGQWKGLGEKDYE 300  
 DB 241 YNLHPLAHARTKGRVYKHLRELAETLEKDLQWRGASVTPGSEIYGQWKGLGEKDYE 300  
 QY 301 IWPBPMTIIMNTRLDKDNDKWLQNGNOELLEFDKYTEALRAHHSYGPQGRGMYLME 360  
 DB 301 IWPBPMTIIMNTRLDKDNDKWLQNGNOELLEFDKYTEALRAHHSYGPQGRGMYLME 360  
 QY 361 SSAGCYLAERLHRELAEMGLDRIAMGKRSMFSGVRYOLYGLATKODLDIFNOSHQK 420  
 DB 361 SSAGCYLAERLHRELAEMGLDRIAMGKRSMFSGVRYOLYGLATKODLDIFNOSHQK 420  
 QY 421 TRLKFEKLSYQEMVYKELRQISEDNQOLNYFNKLSKONKAKYLESLSEIMSKLRTA 480  
 DB 421 TRLKFEKLSYQEMVYKELRQISEDNQOLNYFNKLSKONKAKYLESLSEIMSKLRTA 480  
 QY 481 EENRIVRQRTKMOHONREEMDAHDFPMDSIKQIHERDAKEENFEMLQOQERAKYVQ 540  
 DB 481 EENRIVRQRTKMOHONREEMDAHDFPMDSIKQIHERDAKEENFEMLQOQERAKYVQ 540  
 QY 541 OQONINPSNDCKRKAEEVSSFIPOEKEMEEFYEEBMLIKOQEKQEMDKKHHEI 600  
 DB 541 OQONINPSNDCKRKAEEVSSFIPOEKEMEEFYEEBMLIKOQEKQEMDKKHHEI 600  
 QY 601 FDLKEFDEALBOLMYKGLHNEED 625  
 DB 601 FDLKEFDEALBOLMYKGLHNEED 625  
 RESULT 2  
 AAG91534 standard; protein; 227 AA.  
 ID AAG91534  
 XX AAG91534;  
 AC AAG91534;  
 XX 26-SEP-2001 (first entry)  
 DT 26-SEP-2001 (first entry)  
 XX C glutamicum protein fragment SEQ ID NO: 5288.  
 DE C glutamicum protein fragment SEQ ID NO: 5288.  
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KM organic acid synthesis.  
 XX Corynebacterium glutamicum.  
 OS Corynebacterium glutamicum.  
 XX EPI108790-A2.  
 FN EPI108790-A2.  
 XX 20-JUN-2001.  
 PD 20-JUN-2001.  
 XX 18-DEC-2000; 2000EP-00127688.  
 PF 18-DEC-2000; 2000EP-00127688.  
 XX

PR 16-DEC-1999; 99UP-00377484.  
 PR 07-APR-2000; 2000UP-00159162.  
 PR 03-AUG-2000; 2000UP-00280988.  
 XX  
 PA (KYO) KYOMA HAKKO KOGYO KK.  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX WPI: 2001-376931/40.  
 DR N-PSDB; AAH66753.  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analyzing  
 PT expression profile or pattern of a gene and identifying homologous gene.  
 XX  
 PS Claim 17; SEQ ID NO 5288; 246pp + Sequence Listing; English.  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of Coryneform bacterium, measuring expression amount and analysing  
 CC the expression profile or expression pattern of a gene derived from  
 CC Coryneform bacterium, and identifying a homologue of a gene derived from  
 CC Coryneform bacterium. Coryneform bacteria are useful for producing amino  
 CC acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described in the  
 CC exemplification of the invention. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from the European Patent Office  
 CC  
 SQ Sequence 227 AA;  
 QY Query Match 1.4%; Score 9; DB 4; Length 227;  
 DB Best Local Similarity 100.0%; Pred. No. 2.6;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 170 DSDALDDSD 178  
 DB 179 DSDALDDSD 187  
 RESULT 3  
 AAB79177  
 ID AAB79177 standard; protein; 237 AA.  
 XX AAB79177;  
 AC AAB79177;  
 XX 30-APR-2001 (first entry)  
 DT 30-APR-2001 (first entry)  
 XX Corynebacterium glutamicum HA protein sequence SEQ ID NO:310.  
 DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:310.  
 XX Corynebacterium glutamicum; homeostasis; adaptation; HA protein;  
 KW fine chemical production; organic acid; proteinogenic amino acid;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;  
 KW diagnosis; Corynebacterium diptheriae; genetic engineering;  
 KW Brevibacterium; environmental condition.  
 XX Corynebacterium glutamicum.  
 OS Corynebacterium glutamicum.  
 XX WO200100842-A2.  
 FN WO200100842-A2.  
 XX 04-JAN-2001.  
 PD 04-JAN-2001.  
 XX 23-JUN-2000; 2000WO-IB000911.  
 PF 23-JUN-2000; 2000WO-IB000911.  
 XX 25-JUN-1999; 99US-0141031P.  
 PR 08-JUL-1999; 99DE-01031636.  
 PR 09-JUL-1999; 99DE-01032125.  
 PR 09-JUL-1999; 99DE-01032126.  
 PR 09-JUL-1999; 99DE-01032127.  
 PR 09-JUL-1999; 99DE-01032127.  
 XX

PR 09-JUL-1999; 99DE-01032128.  
 PR 09-JUL-1999; 99DE-01032129.  
 PR 09-JUL-1999; 99DE-01032226.  
 PR 14-JUL-1999; 99DE-01032920.  
 PR 14-JUL-1999; 99DE-01032922.  
 PR 14-JUL-1999; 99DE-01032924.  
 PR 14-JUL-1999; 99DE-01032928.  
 PR 14-JUL-1999; 99DE-01032930.  
 PR 14-JUL-1999; 99DE-01032933.  
 PR 14-JUL-1999; 99DE-01032973.  
 PR 14-JUL-1999; 99DE-01033002.  
 PR 14-JUL-1999; 99DE-01033003.  
 PR 14-JUL-1999; 99DE-01033005.  
 PR 14-JUL-1999; 99DE-01033006.  
 PR 31-AUG-1999; 99DE-01041378.  
 PR 31-AUG-1999; 99DE-01041379.  
 PR 31-AUG-1999; 99DE-01041390.  
 PR 03-SEP-1999; 99DE-01041391.  
 PR 03-SEP-1999; 99DE-01042088.  
 XX (BAD ) BASF AG.  
 XX  
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
 XX  
 DR WPI; 2001-061974/07.  
 DR N-PSDB; AAF71292.  
 XX  
 XX  
 PT New isolated Corynebacterium glutamicum nucleic acid for production or  
 PT modulation of production of fine chemicals such as amino acids,  
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins or  
 PT enzymes.  
 XX  
 PS Claim 20; Page 553-554; 712pp; English.  
 XX  
 XX AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis  
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The C.  
 CC glutamicum HA genes (I) can be used in vectors for expression in host  
 CC cells and production of fine chemicals, such as, an organic acid,  
 CC proinogenic or nonproinogenic amino acid (preferred), purine or  
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated  
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,  
 CC polycyclic or enzyme. The amino acids produced can be lysine, glutamine,  
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,  
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,  
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can  
 CC be modulated. The presence of (I) or HA proteins encoded by then are used  
 CC for diagnosing the presence or activity of Corynebacterium diphtheriae.  
 CC (I) can be used to map the C. glutamicum genome or can be used as markers  
 CC for genetically engineered Corynebacterium or Brevibacterium. The HA  
 CC proteins encoded by the (I) are used to maintain homeostasis in C.  
 CC glutamicum or help the microorganism to adapt to different environmental  
 CC conditions  
 XX  
 SQ Sequence 237 AA;  
 XX  
 Query Match 1.4%; Score 9; DB 4; Length 237;  
 Best Local Similarity 100.0%; Pred. No. 2.7;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 170 DSDALDSD 178  
 Db 189 DSDALDSD 197  
 XX  
 RESULT 4  
 ID AAB18189 standard; protein; 483 AA.  
 XX  
 AC AAB18189;  
 XX  
 DT 07-NOV-2000 (first entry)  
 XX

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO.46.  
 XX  
 XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
 KM antimalarial; malaria; protozoacide; infection; insecticide.  
 XX  
 XX Plasmodium falciparum.  
 OS  
 PN WO200025728-A2.  
 XX  
 PD 11-MAY-2000.  
 XX  
 XX  
 PF 05-NOV-1999; 99WO-US026796.  
 XX  
 PR 05-NOV-1998; 98US-0107131P.  
 XX  
 XX (HOFF/) HOFFMAN S.  
 PA (CARU/) CARUCI D.  
 PA (GARD/) GARDNER M.  
 PA (VENT/) VENTER J C.  
 XX  
 XX  
 PI Hoffman S, Carucci D, Gardner M, Venter JC,  
 XX  
 DR WPI; 2000-365347/31.  
 XX  
 PT Proteins encoded by chromosome 2 of the human malarial parasite,  
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
 PT diagnosis of P.falciparum infection.  
 XX  
 PS Disclosure, Page 112-113; 577pp; English.  
 XX  
 XX The present invention describes proteins and their fragments (I) encoded  
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
 CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
 CC vaccines against P. falciparum infection comprising (I) or (II). (I) and  
 CC (II) are useful for the development of vaccines against P. falciparum  
 CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to  
 CC immunogens comprising the sequences of (I), are useful in the detection  
 CC of infection with P. falciparum. Furthermore, (I) (especially when they  
 CC are rifins or secreted or membrane proteins) can aid the identification  
 CC of drugs to treat or prevent P. falciparum infection, or they can be used  
 CC to identify drug resistance in P. falciparum. Sequencing of the  
 CC Plasmodium chromosome 2 and the subsequent identification of proteins  
 CC encoded by it will help to expand our understanding of parasite biology,  
 CC a process hampered by the complexity of the parasitic lifecycle, and  
 CC provide new targets for vaccine and drug development. Parasite resistance  
 CC to drugs and mosquito resistance to insecticides have led to a resurgence  
 CC of malaria in many parts of the world, and there is a pressing need for  
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352  
 CC represent nucleotide and protein sequences given in the present  
 CC invention, but which are not specifically mentioned within the  
 CC specification  
 XX  
 SQ Sequence 483 AA;  
 XX  
 Query Match 1.3%; Score 8; DB 3; Length 483;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 ISKKNNK 54  
 Db 33 ISKKNNK 40  
 XX  
 RESULT 5  
 ID AEU23489 standard; protein; 916 AA.  
 XX  
 AC AEU23489;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #9016.  
 XX



Query Match 1.3%; Score 8; DB 7; Length 1553;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 RLHRELA 378  
 |||||  
 Db 493 RLHRELA 500

RESULT 7  
 AAE19162  
 ID AAE19162 standard; protein; 1572 AA.  
 XX  
 AC AAE19162;  
 XX  
 DT 21-MAY-2002 (first entry)  
 XX  
 DE Human kinase polypeptide (PKIN-20).  
 XX  
 KW Human: kinase polypeptide; PKIN-20; gene therapy; Addison's disease;  
 KW leukemia; immune disorder; lymphoma; melanoma; developmental disorder;  
 KW acquired immunodeficiency syndrome; AIDS; allergy; anaemia; hyperextension;  
 KW asthma; Crohn's disease; rheumatoid arthritis; bursitis; atherosclerosis;  
 KW cirrhosis; hepatitis; psoriasis; Cushing's syndrome; cytostatic; cancer;  
 KW cholestasis; anti-HIV; cardiovascular disorder; Niemann-Pick's disease;  
 KW lipid disorder; fatty liver; Gaucher's disease; myocardial infarction;  
 KW drug screening; transgenic animal; antiinflammatory; hepatotropic;  
 KW hypotensive; cardiac; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..37  
 FT Protein 38..1572  
 FT Domain /note="Mature human PKIN-20"  
 FT /note="Eukaryotic proetin kinase domain"  
 FT Binding-site 887..935  
 FT /note="Phorbol esters/diacylglycerol binding domain"  
 FT Binding-site 900..963  
 FT /note="Phorbol esters/diacylglycerol binding domain"  
 FT Domain 956..1074  
 FT /note="CNH domain"  
 FT Domain 1100..1380  
 FT /note="CNH domain"  
 XX  
 XX MO200208399-A2.  
 XX  
 XX 31-JAN-2002.  
 XX  
 XX 20-JUL-2001; 2001WO-US023092.  
 XX  
 XX 21-JUL-2000; 2000US-0220038P.  
 XX 28-JUL-2000; 2000US-0222112P.  
 XX 04-AUG-2000; 2000US-0222831P.  
 XX 11-AUG-2000; 2000US-0224729P.  
 XX  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX (THOR/) THORNTON M.  
 XX  
 XX Thornton M, Yue H, Khan FA, Gururajan R, Hafalia AD, Walla NK,  
 XX Patterson C, Ramkumar J, Gandhi AR, Policky JL, Baughn MR,  
 XX Tribouley CM, Bandman O, Nguyen DB, Lu Y, Buford N, Lal P, Ding L,  
 XX Yao MG, Elliott VS, Reclipon SA, Kearney L, Lu DM, Greenwald SR,  
 XX Tang YT, Xu Y, Walsh RT, Gietzen KU, Yang J, Hillman JL;  
 XX  
 XX WPI; 2002-206083/26.  
 XX N-PSDB; AAD30567.  
 XX  
 XX New human kinase polypeptide, useful in diagnosis, prevention and  
 XX treatment of cancer, immune disorder, growth and developmental disorder,  
 XX cardiovascular disorder and lipid disorder.

XX  
 PS Claim 1; Page 170-174; 196pp; English.  
 XX  
 CC The present invention relates to an isolated human kinase polypeptide  
 CC (PKIN) or a biologically active/immunogenic fragment of PKIN. PKIN is  
 CC useful for diagnosing, treating and preventing cancer (e.g., leukaemia,  
 CC lymphoma, melanoma), an immune disorder (e.g., acquired immunodeficiency  
 CC syndrome (AIDS), Addison's disease, allergy, anaemia, asthma, Crohn's  
 CC disease, rheumatoid arthritis), a growth and developmental disorder (e.g.,  
 CC bursitis, cirrhosis, hepatitis, psoriasis, Cushing's syndrome), a  
 CC cardiovascular disorder (e.g., atherosclerosis, hypertension, myocardial  
 CC infarction), and a lipid disorder (e.g., fatty liver, cholestasis,  
 CC Gaucher's disease, Niemann-Pick's disease). PKIN is useful in a number of  
 CC drug screening techniques and to analyse the proteome of a tissue or cell  
 CC type. PKIN is useful for creating knockin humanised animals or transgenic  
 CC animals to model human diseases, in somatic or germ-line gene therapy, to  
 CC generate a transcript image of a tissue or cell type, for detecting  
 CC differences in the chromosomal location due to translocation, inversion,  
 CC etc., among normal, carrier or affected individuals, and as hybridisation  
 CC probes for mapping naturally occurring genomic sequences. PKIN is useful  
 CC in southern or northern analysis, dot blot or other membrane-based  
 CC technologies, in PCR technologies, in dipstick, pin, multiformat enzyme  
 CC linked immunosorbent (ELISA)-like assays and in microarrays utilising  
 CC fluids or tissues from patients to detect altered PKIN expression. The  
 CC present sequence is human PKIN-20  
 XX  
 SO Sequence 1572 AA;

QY 371 RLHRELA 378  
 |||||  
 Db 493 RLHRELA 500

RESULT 8  
 AAG71491  
 ID AAG71491 standard; peptide; 15 AA.  
 XX  
 AC AAG71491;  
 XX  
 DT 28-FEB-2003 (first entry)  
 XX  
 DE Prostatic specific membrane antibody protein 11.88 N-terminal peptide.  
 XX  
 XX Human; prostatic specific membrane antibody protein 11.88; antibody;  
 XX prostatic cancer; prostatic benign tumour; folic acid; metabolism;  
 XX haemopathy; human immunodeficiency virus; HIV; immunological disease;  
 XX inflammation.  
 XX  
 XX Homo sapiens.  
 XX  
 XX CN1352107-A.  
 XX  
 XX 05-JUN-2002.  
 XX  
 XX 06-NOV-2000; 2000CN-00127237.  
 XX  
 XX 06-NOV-2000; 2000CN-00127237.  
 XX  
 XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
 XX  
 XX Mao Y, Xie Y;  
 XX  
 XX WPI; 2002-692403/75.  
 XX  
 XX New human prostatic specific membrane antibody protein 11.88 polypeptide  
 XX for treating e.g. prostatic cancer, folic acid metabolic fault, human  
 XX immunodeficiency virus infection, immunological diseases, and  
 XX inflammations.



PS Example 5; Page 19 (disclosure); 33pp; Chinese.

XX The invention relates to the human prostatic specific membrane antibody

CC protein 11.88, a polynucleotide encoding the polypeptide and a DNA

CC recombinant process used to produce the polypeptide. The polypeptide

CC and the polynucleotide are used for treating various diseases, such as

CC prostatic cancer, prostatic benign tumour, other tumours, folic acid

CC metabolism fault, haemopathy, human immunodeficiency virus (HIV)

CC infection, immunological diseases and inflammations. This sequence

CC represents a human prostatic specific membrane antibody protein 11.88 N-terminal peptide, used in ELISA

XX Sequence 15 AA:

SQ

Query Match 1.1%; Score 7; DB 5; Length 15;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 SKKNKK 54

Db 7 SKKNKK 13

RESULT 9

ABP82841

ID ABP82841 standard; peptide; 18 AA.

XX

AC ABP82841;

XX

DT 04-MAR-2003 (first entry)

XX

DE G protein-coupled receptor (GPCR) antigenic peptide SEQ ID NO:1514.

XX

KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;

XX G protein-coupled receptor modulator; antibody; immune-related disease;

KW growth-related disease; cell regeneration-related disease; AIDS; cancer;

KW immunological-related cell proliferative disease; autoimmune disease;

KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;

KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;

KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;

KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;

KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;

KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;

KW ulcer.

XX

XX Homo sapiens.

OS

XX

PN WO200261087-A2.

XX

PD 08-AUG-2002.

XX

PF 19-DEC-2001; 2001WO-US050107.

XX

PR 19-DEC-2000; 2000US-0257144P.

XX

PA (LIFE-) LIFESPAN BIOSCIENCES INC.

XX

PI Burmer GC, Roush CL, Brown JP;

XX

DR WPI; 2003-046718/04.

XX

PT New isolated antigenic peptides e.g., for G protein-coupled receptors

PT (GPCR), useful for diagnosing and designing drugs for treating conditions

PT in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or

PT autoimmune diseases.

XX

PS Claim 1; Fig 2; 523pp; English.

XX

CC The present invention describes antigenic peptides (1) comprising: (a)

CC any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino

CC acids. Also described: (1) an assay for the detection of a particular G

CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;

CC and (2) an isolated antibody having high specificity and high affinity or

CC avidity for a particular GPCR. (1) can be used as GPCR modulators and in

CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an

CC antibody against a particular GPCR, and in the production of specific

CC antibodies. The peptides and antibodies are also useful for detecting the

CC presence or absence of corresponding GPCRs. The antigenic peptides for

CC GPCRs and antibodies are useful for diagnosing and designing drugs for

CC treating immune-related diseases, growth-related diseases, cell

CC regeneration-related disease, immunological-related cell proliferative

CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,

CC atherosclerosis, bacterial, fungal, protozoan or viral infections,

CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute

CC inflammation, allergies, Crohn's disease, diabetes, graft versus host

CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,

CC anxiety, depression, schizophrenia, dementia, mental retardation, memory

CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,

CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or

CC any other disorder in which GPCRs are involved. The antibodies may be

CC used in immunoassays and immunodiagnosis. ABP2523 to ABP24269 encode

CC GPCR proteins given in ABP1675 to ABP82018, which are used in the

CC exemplification of the present invention

XX

SQ

Query Match 1.1%; Score 7; DB 6; Length 18;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 DSDDDLA 182

Db 3 DSDDDLA 9

RESULT 10

AAE12564

ID AAE12564 standard; peptide; 32 AA.

XX

AC AAE12564;

XX

DT 15-JUN-2002 (first entry)

XX

DE p35srj/CITBD2 p300-CH1 interacting domain.

XX

KW Human; transcription transactivator protein; CREB-binding protein;

KW CITED family; CBP/p300 interacting transactivators with BD-rich tails;

KW HCITEDX protein; hypoxia signalling; inflammation; tumour angiogenesis;

KW cholesterol biosynthesis; ischaemic heart disease; Neheron's disease;

KW angiogenesis; atopic disease; cytostatic; vasotropic; gene therapy;

KW p35srj/CITBD2 p300-CH1 interacting domain.

XX

XX Homo sapiens.

OS

XX

PN WO200168846-A2.

XX

PD 20-SEP-2001.

XX

PF 16-MAR-2001; 2001WO-GB001201.

XX

PR 17-MAR-2000; 2000GB-0006572.

XX

PA (ISIS-) ISIS INNOVATION LTD.

XX

PI Bhattacharya S, Braganca J, Swingle T;

XX

DR WPI; 2001-596907/67.

XX

PT New transcription transactivator protein of CITED family, termed HCITEDX,

PT useful for controlling hypoxia signaling, inflammation, activating

PT cholesterol uptake genes and identifying compounds interfering with

PT HCITEDX function.

XX

PS Example 1; Page 50; 95pp; English.

XX

CC The patent discloses transcription transactivator protein, which is a

CC member of CITED (CREB-binding protein (CBP)/p300 interacting  
 CC transactivators with ED-rich tails) family, designated as HCITEDX protein  
 CC and its corresponding polynucleotides. HCITEDX proteins are useful for  
 CC identifying compounds which modulate the cytoplasmic sequestration of the  
 CC HCITEDX protein. They are useful in the control of hypoxia signalling and  
 CC in the activation of genes involved in cholesterol uptake, cholesterol  
 CC biosynthesis and in the control of inflammation. They are therapeutically  
 CC useful in number of different indications, particularly reduction of  
 CC cholesterol biosynthesis and inflammation, prevention of tumour  
 CC angiogenesis and treatment of any other condition. HCITEDX proteins are  
 CC useful in screening methods for identification of compounds which  
 CC interfere with the function of HCITEDX and useful for modulating hypoxia  
 CC signalling, tumour angiogenesis or cholesterol synthesis. HCITEDX  
 CC modulators are useful for preventing tumour angiogenesis and treating of  
 CC ischaemic heart disease. Compounds which promote the translocation of  
 CC cytoplasmic HCITEDX in breast cancer cells are useful as leads in the  
 CC development of anti-breast cancer drugs and compounds which prevent the  
 CC translocation of HCITEDX into the nucleus enhance angiogenesis. HCITEDX  
 CC modulators are useful for treating Netherton's disease or atopic disease.  
 CC HCITEDX sequences are also used in gene therapy. The present peptide  
 CC sequence is the p300-CHI interacting domain of p35arf/CITED2. This  
 CC sequence is used to clone full length HCITEDX

SQ Sequence 32 AA;

Query Match 1.1%; Score 7; DB 4; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 EMGLDRI 384  
 DB 11 EMGLDRI 17

RESULT 11  
 ID AAO04467 standard; protein; 34 AA.

AAO04467;

06-NOV-2001 (first entry)

Human polypeptide SEQ ID NO 18359.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;

vacine; peptide therapy; stem cell growth factor; haematopoiesis;

tissue growth factor; immunomodulatory; cancer; leukaemia;

nervous system disorders; arthritis; inflammation.

Homo sapiens.

WO200164835-A2.

07-SEP-2001.

26-FEB-2001; 2001WO-US004927.

28-FEB-2000; 2000US-00515126.

18-MAY-2000; 2000US-00577409.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-514838/56.

N-PADB; AA184398.

Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
 and treating e.g. leukemia, inflammation and immune disorders.

Claim 20, SEQ ID NO 18359; 1399pp + Sequence Listing; English.

The invention relates to human polynucleotides (AA19941-AA193841) and

CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, hematopoietic regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

SQ Sequence 34 AA;

Query Match 1.1%; Score 7; DB 4; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 SKKNKK 54  
 DB 23 SKKNKK 29

RESULT 12  
 ID AAB13241 standard; peptide; 36 AA.

AAB13241;

11-JAN-2001 (first entry)

Caenorhabditis elegans conserved sequence #5.

Caenorhabditis elegans; metabolic enzyme; AKT kinase; daf-18;

insulin signalling pathway; daf-2; age-1; insulin receptor; pi 3-kinase;

PKB kinase; PTEN lipid phosphatase; antidiabetic; anorectic; obesity;

diabetes.

Caenorhabditis elegans.

WO200033068-A1.

08-JUN-2000.

02-DEC-1999; 99WO-US028529.

03-DEC-1998; 98US-00205658.

(GENO) GEN HOSPITAL CORP.

Ruvkun G, Ogg S;

WPI; 2000-423022/36.

Diagnosing and treating obesity and impaired glucose tolerance using  
 modulators of daf-18 expression and/or activity.

Disclosure; Page 167; 402pp; English.

The present sequence is conserved protein region of a Caenorhabditis  
 CC elegans homologue of a key metabolic enzyme. A number of C. elegans genes  
 CC have been identified as homologues of genes in the mammalian insulin  
 CC signalling pathway. The C. elegans age-1 gene encodes a homologue of the  
 CC mammalian pi 3-kinase whilst daf-2 encodes a homologue of the mammalian  
 CC insulin receptor. The C. elegans AKT kinase and PKB kinase act downstream  
 CC of daf-2 and age-1, just as their mammalian homologues act downstream of  
 CC insulin signalling. The C. elegans PTEN lipid phosphatase homologue, DAF-  
 CC 18, has been found to act upstream of AKT in the pathway. This discovery  
 CC has enabled mammalian PTEN action to be mapped to the insulin signalling  
 CC pathway. Conserved DAF motifs can be used to design probes to identify  
 CC mammalian DAF homologues and thus to identify individuals with a  
 CC predisposition toward the development of glucose intolerance conditions,

CC such as obesity and diabetes  
XX  
SQ Sequence 36 AA;

Query Match 1.1%; Score 7; DB 3; Length 36;  
Best Local Similarity 100.0%; Pred. No. 62;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 LKQISD 444  
DB 2 LKQISD 8

## RESULT 13

AA018208 standard; peptide; 43 AA.

XX  
AC AA018208;

DT 13-SEP-2002 (first entry)

XX Murine rotavirus non-structural protein 4 NSP4 fragment #2.

XX Rotavirus; non-structural protein 4; NSP4; protein coordinate data;  
KM crystal structure; antiviral; vaccine; diarrhoea; gastroenteritis.

XX Murine rotavirus.

XX WO200244330-A2.

XX 06-JUN-2002.

XX 28-NOV-2001; 2001WO-US044555.

XX 01-DEC-2000; 2000US-0250810P.

XX (AMHP) AMERICAN HOME PROD CORP.

XX Bowman GD, Levy O, Nodelman I, Schutt CE, Zamb TJ;

XX WPI; 2002-537448/57.

XX Novel crystal of oligomerization domain of rotavirus nonstructural  
PT protein 4, useful for determining 3 dimensional crystal structure of the  
PT domain that is used for identifying agents which interact with the  
PT domain.

XX Disclosure; Fig 1B; 81pp; English.

XX The present invention provides the crystal structure of the  
CC oligomerization domain of the rotavirus non-structural protein 4 (NSP4).

XX Rotavirus is a major cause of gastroenteritis, and the structure can be  
CC used to construct peptides useful in vaccines to protect against  
CC infection. The present sequence is a fragment of a rotavirus NSP4 protein

XX Sequence 43 AA;

SQ

Query Match 1.1%; Score 7; DB 5; Length 43;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 VVKELRQ 440  
DB 8 VVKELRQ 14

## RESULT 14

AA018207 standard; peptide; 43 AA.

XX  
AC AA018207;

DT 13-SEP-2002 (first entry)

XX Murine rotavirus non-structural protein 4 NSP4 fragment #1.

XX Rotavirus; non-structural protein 4; NSP4; protein coordinate data;  
KM crystal structure; antiviral; vaccine; diarrhoea; gastroenteritis.

XX Murine rotavirus.

XX WO200244330-A2.

XX 06-JUN-2002.

XX 28-NOV-2001; 2001WO-US044555.

XX 01-DEC-2000; 2000US-0250810P.

XX (AMHP) AMERICAN HOME PROD CORP.

XX Bowman GD, Levy O, Nodelman I, Schutt CE, Zamb TJ;

XX WPI; 2002-537448/57.

XX Novel crystal of oligomerization domain of rotavirus nonstructural  
PT protein 4, useful for determining 3 dimensional crystal structure of the  
PT domain that is used for identifying agents which interact with the  
PT domain.

XX Disclosure; Fig 1B; 81pp; English.

XX The present invention provides the crystal structure of the  
CC oligomerization domain of the rotavirus non-structural protein 4 (NSP4).

XX Rotavirus is a major cause of gastroenteritis, and the structure can be  
CC used to construct peptides useful in vaccines to protect against  
CC infection. The present sequence is a fragment of a rotavirus NSP4 protein

XX Sequence 43 AA;

SQ

Query Match 1.1%; Score 7; DB 5; Length 43;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 VVKELRQ 440  
DB 8 VVKELRQ 14

## RESULT 15

AAU43846 standard; protein; 51 AA.

XX AAU43846;

DT 13-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #4742.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KM dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

XX 02-JUN-2000; 2000US-0208841P.

XX 07-JUL-2000; 2000US-0216747P.

PA (CORI-) CORIXA CORP.

XX Skeily YAW, Persing DH, Mitcham JL, Wang SS, Bhactia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX

DR WPI; 2001-616774/71.  
DR N-PSDB; AAS59521.

XX  
XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.

PS Example 1; SEQ ID NO 5041; 1069pp; English.

XX  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
SQ Sequence 51 AA;

Query Match 1.1%; Score 7; DB 4; Length 51;

Best Local Similarity 100.0%; Pred. No. 85;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 SGRGN 99  
Db 34 SGRGN 40

Search completed: April 6, 2004, 19:42:00  
Job time : 62 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:39:52 ; Search time 20 Seconds  
(without alignments)  
3005.985 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 625  
Sequence: 1 MSSRAGPMSEKKNVQGGYRP.....EFDEALQQLMYKHGLHNEDD 625

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: PIR 78: \*  
2: pir1: \*  
3: pir2: \*  
4: pir3: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.4	603	2 T27901	hypothetical prote
2	8	1.3	353	2 T33782	hypothetical prote
3	8	1.3	483	2 F71619	hypothetical prote
4	8	1.3	586	2 T19406	hypothetical prote
5	7	1.1	69	2 G84303	snRNP homolog (imp
6	7	1.1	101	2 D84164	hypothetical prote
7	7	1.1	135	2 B70456	thioredoxin - Aqu
8	7	1.1	141	2 C97338	hydroxymyristoyl-(
9	7	1.1	145	2 A12869	NTP pyrophosphohyd
10	7	1.1	147	1 S24310	flavodoxin - Desul
11	7	1.1	147	2 C97646	hypothetical prote
12	7	1.1	148	1 OTBY6	cytochrome-c oxida
13	7	1.1	149	2 D69831	conserved hypotet
14	7	1.1	152	2 G96010	hypothetical expor
15	7	1.1	182	2 T15118	hypothetical prote
16	7	1.1	207	2 T49552	hypothetical prote
17	7	1.1	212	2 C84007	negative regulator
18	7	1.1	215	2 S77663	multicopy phage re
19	7	1.1	253	2 H72384	conserved hypotet
20	7	1.1	257	2 C40304	neurotrophin-3 pre
21	7	1.1	257	2 F50400	neurotrophin-3 pre
22	7	1.1	257	2 F75476	probable uroporphyr
23	7	1.1	258	2 S09155	neurotrophin-3 pre
24	7	1.1	268	2 H72483	hypothetical prote
25	7	1.1	271	2 I39491	hypothetical prote
26	7	1.1	272	2 A71710	hypothetical prote
27	7	1.1	275	2 H87546	enoyl-CoA hydratase
28	7	1.1	277	2 G71822	crRNA delta(2)-18op
29	7	1.1	277	2 E97703	hypothetical prote

30	7	1.1	278	2 T30450	late expression fa
31	7	1.1	279	2 S67670	hypothetical prote
32	7	1.1	282	2 A35781	hippocampus-derive
33	7	1.1	283	2 F85700	hypothetical prote
34	7	1.1	283	2 A90843	ychb protein [semi
35	7	1.1	283	2 B47706	ychb protein - Bac
36	7	1.1	285	2 T41656	vacuolar ATP synth
37	7	1.1	286	2 B41810	transcription fact
38	7	1.1	286	2 B82564	acetylxylian estera
39	7	1.1	301	2 JS0681	hypothetical prote
40	7	1.1	301	2 B85087	probable phosphogl
41	7	1.1	304	2 D82304	glutamate-tRNA synt
42	7	1.1	309	2 F97168	nucleoside-diphosp
43	7	1.1	320	2 S69547	transcription init
44	7	1.1	321	2 AH0412	probable glutamyl-
45	7	1.1	324	2 T00939	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

T27901 hypothetical protein ZK546.13 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #ext\_change 15-Oct-1999

C:Accession: T27901

R:Hallsworth, K. Submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid ZK546.

A:Reference number: Z20437

A:Accession: T27901

A>Status: preliminary; translated from GB/EMBL/DBDJB

A:Molecule type: DNA

A:Residues: 1-603 <HAL>

A:Cross-references: EMBL:U29380; PIDN:AAA68747.1; CESP:ZK546.13

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:ZK546.13

A:Introns: 26/3; 111/3; 196/3; 275/3; 331/1; 362/3; 438/3; 500/3

##### Query Match

Best Local Similarity 1.4%; Score 9; DB 2; Length 603;  
Matches 9; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

QY	469	LEIMSEKLR 477
DB	433	LEIMSEKLR 441

##### RESULT 2

T33782 hypothetical protein C39F7.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #ext\_change 04-Mar-2000

C:Accession: T33782

R:Maggi, L.; Scheet, P.; Dubbelde, C. Submitted to the EMBL Data Library, October 1998

A:Description: The sequence of C. elegans cosmid C39F7.

A:Reference number: Z21407

A:Accession: T33782

A>Status: preliminary; translated from GB/EMBL/DBDJB

A:Molecule type: DNA

A:Residues: 1-353 <MAG>

A:Cross-references: EMBL:AF101310; PIDN:AAC69214.1; GSPDB:GN00023; CESP:C39F7.5

A:Experimental source: strain Bristol N2; clone C39F7

C:Genetics:

A:Gene: CESP:C39F7.5

A:Map position: 5

A:Introns: 14/2; 45/3; 224/1

C:Superfamily: Caenorhabditis elegans hypothetical protein C39F7.5

Query Match 1.3%; Score 8; DB 2; Length 353;

Best Local Similarity 100.0%; Pred. No. 10;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 21 EYEOLVOG 28  
|||||  
342 EYEOLVOG 349

## RESULT 3

F71619  
hypothetical protein PFB0235w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: F71619  
C:Author: M.J.; Tetteelin, H.; Canucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
R.Gardner, M.J.; Salzman, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: F71619  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-483 <GAR>  
A:Cross-references: GB:AE001382; GB:AE001362; NID:93845130; PIDN:AACT1836.1; PID:9384513  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0235w

Query Match 1.3%; Score 8; DB 2; Length 483;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ISKKNKK 54  
|||||  
DB 33 ISKKNKK 40

## RESULT 4

t19406  
hypothetical protein C18E9.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T19406  
R:Sim, M.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z19120  
A:Accession: T19406  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-586 <WIL>  
A:Cross-references: EMBL:Z70034; PIDN:CAA93857.1; GSPDB:GN00020; CESP:C18E9.8  
A:Experimental source: clone C18E9  
C:Genetics:  
A:Gene: CESP:C18E9.8  
A:Map position: 2  
A:introns: 15873; 2693/3; 3543/3; 4932/2; 538/2  
C:Superfamily: Caenorhabditis elegans hypothetical protein C18E9.8

Query Match 1.3%; Score 8; DB 2; Length 586;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 RELAEVLE 269  
|||||  
DB 495 RELAEVLE 502

## RESULT 5

G84303  
snRNP homolog (imported) - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: G84303

R;N; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.  
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Lie  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: G84303  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-69 <STO>  
A:Cross-references: GB:AE004437; NID:910580991; PIDN:AAI19795.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: snp

Query Match 1.1%; Score 7; DB 2; Length 69;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 464 VLRESLE 470  
|||||  
DB 17 VLRESLE 23

## RESULT 6

D84164  
hypothetical protein Vng0041c (imported) - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: D84164  
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.  
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Lie  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: D84164  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-101 <STO>  
A:Cross-references: GB:AE004437; NID:910579692; PIDN:AAI18680.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG0041C  
C:Superfamily: Halobacterium plasmid pNRC100 hypothetical protein H0011

Query Match 1.1%; Score 7; DB 2; Length 101;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 AHARTKG 254  
|||||  
DB 55 AHARTKG 61

## RESULT 7

B70456  
thioredoxin - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C:Accession: B70456  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ove  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: B70456  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-135 <AGP>  
A:Cross-references: GB:AE000757; NID:92984092; PIDN:AACT07635.1; PID:92984097; GB:AE00065  
A:Experimental source: strain VFS  
C:Genetics:

A:Gene: trxA2

Query Match 1.1%; Score 7; DB 2; Length 135;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 206 KFFGSID 212  
Db 109 KFFGSID 115

RESULT 8

C97338

hydroxymyristoyl-(acyl carrier protein) dehydratase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C&gt;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001

C:Accession: C97338

R:Colling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A&gt;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C1C

A:Reference number: A96900; MWID:21359325; PMID:21359325

A:Accession: C97338

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-141 &lt;KUR&gt;

A:Cross-references: GB:AE001437; PIDN:AAK81494.1; PID:G1502667; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

A:Gene: CAC3571

C:Superfamily: (3R)-hydroxymyristoyl-(acyl carrier protein) dehydratase

Query Match 1.1%; Score 7; DB 2; Length 141;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 213 SLSTEQI 219  
Db 2 SLSTEQI 8

RESULT 9

A12869

NTP pyrophosphohydrolase, MutT family [imported] - Agrobacterium tumefaciens (strain C58

C:Species: Agrobacterium tumefaciens

C&gt;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C:Accession: A12869

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Giller, W.; Grant, C.; Guenther, D.; Kutayvln, T.; Levy, R.; Li, M.; McClall

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kreespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A&gt;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MWID:21608550; PMID:11743193

A:Accession: A12869

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-145 &lt;KUR&gt;

A:Cross-references: GB:AE008688; PIDN:AA143375.1; PID:G17740871; GSPDB:GN00186

A:Experimental source: strain C58 (Dugont)

C:Genetics:

A:Gene: Atu2387

A:Map position: circular chromosome

Query Match 1.1%; Score 7; DB 2; Length 145;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 260 LHRELAE 266  
Db 51 LHRELAE 57

RESULT 10

S24310

flavodoxin - Desulfovibrio gigas (ATCC 29494)

C:Species: Desulfovibrio gigas

C&gt;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 05-May-2000

C:Accession: S24310

R:Helms, L.R.; Swenson, R.P.

Biochim. Biophys. Acta 1131, 325-328, 1992

A&gt;Title: The primary structures of the flavodoxins from two strains of Desulfovibrio giga

A:Reference number: S24310; MWID:92329549; PMID:1627649

A:Accession: S24310

A:Molecule type: DNA

A:Residues: 1-147 &lt;HEL&gt;

A:Cross-references: EMBL:664765; NID:940798; PIDN:CAA46012.1; PID:G40799

A:Experimental source: strain ATCC 29494

C:Superfamily: flavodoxin; flavodoxin homology

C:Keywords: electron transfer; flavoprotein; FMN

F/6-143/Domain: flavodoxin homology &lt;FLX&gt;

Query Match 1.1%; Score 7; DB 1; Length 147;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 462 AKVLEES 468  
Db 23 AKVLEES 29

RESULT 11

C97646

hypothetical protein AGR\_C 4330 [imported] - Agrobacterium tumefaciens (strain C58, Cerec

C:Species: Agrobacterium tumefaciens

C&gt;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002

C:Accession: C97646

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A&gt;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume

A:Reference number: A97359; MWID:21608551; PMID:11743194

A:Accession: C97646

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-147 &lt;KUR&gt;

A:Cross-references: GB:AE007869; PIDN:AAK8124.1; PID:G15157558; GSPDB:GN00169

C:Genetics:

A:Gene: AGR\_C 4330

A:Map position: circular chromosome

Query Match 1.1%; Score 7; DB 2; Length 147;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 260 LHRELAE 266  
Db 53 LHRELAE 59

RESULT 12

O7EY6

cytochrome-c oxidase (EC 1.9.3.1) chain VI precursor - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein H6179.2; protein YHR051w

C:Species: Saccharomyces cerevisiae

C&gt;Date: 13-Jun-1983 #sequence\_revision 31-Dec-1992 #text\_change 23-Mar-2001

C:Accession: A22853; S46730; A00494; S48874

R:Wright, R.M.; Ko, C.; Cumsky, M.G.; Poyton, R.O.

J. Biol. Chem. 259, 15401-15407, 1984

A&gt;Title: Isolation and sequence of the structural gene for cytochrome c oxidase subunit I

A:Reference number: A92466; MWID:85080033; PMID:6210289

A:Accession: A22853

A:Molecule type: DNA

A:Residues: 1-148 &lt;WRI&gt;

A:Cross-references: EMBL:M10138; NID:G171294; PIDN:AAA66900.1; PID:G171295



R.Du, Z.  
submitted to the EMBL Data Library, May 1994  
A:Description: The sequence of S. cerevisiae cosmid 8179.  
A:Reference number: S46730  
A:Accession: S46730  
A:Molecule type: DNA  
A:Residues: 1-148 <DUZ>  
A:Cross-references: EMBL:U00062; NID:G488162; PIDN:AA68899.1; PID:G488164; GSPDB:GN0000  
R:Gregor, I.; Tsubota, A. 1982  
J. Biol. Chem. 257, 13081-13087, 1982  
A:Title: The amino acid sequence of cytochrome c oxidase subunit VI from Saccharomyces c  
A:Reference number: A00494; WUID:83030850; PMID:6290493  
A:Accession: A00494  
A:Molecule type: protein  
A:Residues: 41-148 <GRE>  
R:Wright, R.M.; Rosenzweig, B.; Peyton, R.O.  
Nucleic Acids Res. 17, 1103-1120, 1989  
A:Title: Organization and expression of the COX6 genetic locus in Saccharomyces cerevisia  
A:Reference number: S48872; WUID:89160242; PMID:2537949  
A:Accession: S48872  
A:Molecule type: DNA  
A:Residues: 1-23 <MRN>  
A:Cross-references: EMBL:X14452; NID:G3573; PIDN:CAA32622.1; PID:G3576  
C:Comment: Cytochrome-c oxidase is the terminal component of the respiratory chain; it c  
C:Genetics:  
A:Gene: SGD:COX6; MIPS:YHR051W  
A:Cross-references: SGD:S0001093; MIPS:YHR051W  
A:Map position: 8R  
A:Genome: nuclear  
C:Function:  
A:Description: oxidoreductase  
A:Pathway: oxidative phosphorylation; respiratory chain  
C:Superfamily: mammalian cytochrome-c oxidase chain Va  
C:Keywords: membrane-associated complex; mitochondrial inner membrane; mitochondrion; ox  
F:1-40/Domain: transit peptide (mitochondrion) #status predicted <TMP>  
F:41-148/Product: cytochrome-c oxidase chain VI #status experimental <MAT>

Query Match 1.1%; Score 7; DB 1; Length 148;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 EKEFDEA 610  
DB 57 EKEFDEA 63

RESULT 13  
D69831  
conserved hypothetical protein ynfO - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: D69831  
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Biron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chn  
C.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati, B.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler  
A:Authors: Foulger, D.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
A:Authors: Foulger, D.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Foulger, D.; Krogh, S.; Krogh, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
A:Authors: Foulger, D.; Lazarevic, V.; Lee, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
A:Authors: Foulger, D.; Oudega, B.; Park, S.H.; Parro, V.; Sadate, Y.; Sato, T.; Scanlon,  
A:Authors: Foulger, D.; Roche, B.; Rose, M.; Sadate, Y.; Sekowska, A.; Serot  
A:Authors: Foulger, D.; Tanaka, T.; Terpetre, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
A:Authors: Foulger, D.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, B.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; WUID:98044033; PMID:9384377  
A:Accession: A69580  
A:Molecule type: DNA  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Residues: 1-149 <KUN>  
A:Cross-references: GB:Z99109; GB:AL009126; NID:G2633260; PIDN:CAM12871.1; PID:G2633367

A:Experimental source: strain 168  
C:Genetics:  
A:Gene: ynfO  
C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ1207

Query Match 1.1%; Score 7; DB 2; Length 149;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 HARTKGA 255  
DB 87 HARTKGA 93

RESULT 14  
G96010  
hypothetical exported protein, glycine-rich [imported] - Sinorhizobium meliloti (strain )  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: G96010  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo  
A:Reference number: A95842; WUID:21396508; PMID:11481431  
A:Accession: G96010  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-152 <KIR>  
A:Cross-references: GB:AL591985; PIDN:CA049751.1; PID:G15141238; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymb  
R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pla, D.; Chain, P.; Cowle, A.; Davis, R.W.; Dreano, S.; Federpiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.M.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
A:Authors: Kahn, D.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K (hebaul  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; WUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: Smb21521  
A:Genome: plasmid

Query Match 1.1%; Score 7; DB 2; Length 152;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 GNGNGRG 102  
DB 70 GNGNGRG 76

RESULT 15  
T1518  
hypothetical protein T23C6.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 04-Mar-2000  
C:Accession: T1518  
R.Du, Z.  
submitted to the EMBL Data Library, April 1997  
A:Description: The sequence of C. elegans cosmid T23C6.  
A:Reference number: Z18295  
A:Accession: T1518  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-182 <DUZ>  
A:Cross-references: EMBL:AF00091; NID:G1946976; PID:G1946981; PIDN:AA52885.1; GSPDB:GN  
A:Experimental source: strain Bristol N2; clone T23C6  
C:Genetics:  
A:Gene: CESP:T23C6.4  
A:Map position: X  
A:Insertions: 37/1; 90/1; 142/2  
C:Superfamily: Caenorhabditis elegans hypothetical protein T23C6.4

Query Match 1.1%; Score 7; DB 2; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 GRGNNG 100  
 |||||  
 Db 173 GRGNNG 179

Search completed: April 6, 2004, 19:44:01  
 Job time : 22 secs

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:36:06 ; Search time 18 Seconds

(without alignments)  
1807.991 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 625  
Sequence: 1 MSSRAGPMSEKKNVQGGYRP.....EFDEALQMLYKHLNEDD 625

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.4	227	1	NUSB CORGL
2	8	1.3	113	1	RBFA_OCEIH
3	7	1.1	147	1	FABZ_CLOAB
4	7	1.1	147	1	FLAW_DESGI
5	7	1.1	148	1	COX6_YEAST
6	7	1.1	153	1	G6PD_SARBU
7	7	1.1	154	1	NT3_CEREL
8	7	1.1	242	1	Y370_THEMA
9	7	1.1	257	1	NT3_CHICK
10	7	1.1	257	1	NT3_PELCA
11	7	1.1	257	1	NT3_HUMAN
12	7	1.1	258	1	NT3_MOUSE
13	7	1.1	258	1	NT3_RAT
14	7	1.1	269	1	CIT2_MOUSE
15	7	1.1	270	1	CIT2_HUMAN
16	7	1.1	271	1	YNR3_AZOB
17	7	1.1	283	1	ISPE_ECOL6
18	7	1.1	283	1	ISPE_ECOLI
19	7	1.1	285	1	VATD_SCHPO
20	7	1.1	286	1	TOAI_YEAST
21	7	1.1	301	1	YGL4_BACST
22	7	1.1	308	1	CG17_YEAST
23	7	1.1	311	1	MTAA_HELPY
24	7	1.1	312	1	MTAA_HELPY
25	7	1.1	347	1	ITP2_HUMAN
26	7	1.1	360	1	SYFA_AGRIS
27	7	1.1	360	1	SYFA_RHIME
28	7	1.1	372	1	ALR_SYNY3
29	7	1.1	372	1	ISP6_VIBPA
30	7	1.1	372	1	ISP6_VIBVU
31	7	1.1	390	1	TAL_PROMM
32	7	1.1	399	1	T7S1_HUMAN
33	7	1.1	420	1	TRPB_HELPAP

34	7	1.1	453	1	SR54_ECOLI	P07019	escherichia
35	7	1.1	461	1	ARLY_SYNY3	P73257	synchocyst
36	7	1.1	476	1	Y32_YEAST	P53270	saccharomyc
37	7	1.1	483	1	APW3_YEAST	P38153	saccharomyc
38	7	1.1	510	1	EMRB_HAEIN	P44927	haemophilus
39	7	1.1	523	1	UDBG_RABIT	019103	oryctolagus
40	7	1.1	527	1	UDAI_RAT	P36510	rattus norv
41	7	1.1	528	1	UDB4_HUMAN	P36513	homo sapien
42	7	1.1	528	1	UDBA_HUMAN	P36537	homo sapien
43	7	1.1	528	1	UDBA_MACFA	09RT55	macaca fasc
44	7	1.1	529	1	UDB3_MACFA	002663	macaca fasc
45	7	1.1	529	1	UDBB_HUMAN	075310	homo sapien

#### ALIGNMENTS

```

RESULT 1
NUSB CORGL          STANDARD;          PRT;          227 AA.
ID  NUSB CORGL
AC  08N033;
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  N utilization substance protein B homolog (Nusb protein).
GN  NUSB OR GGL1618.
OS  Corynebacterium glutamicum (Brevibacterium flavum).
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX  NCBI_TaxID=1718;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA  Nakagawa S.;
RT  "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL  Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: Involved in the transcription termination process (By
CC  similarity).
CC  -1- SIMILARITY: Belongs to the nusb family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AP005279; BAB99011.1; -.
DR  HAMAP; MF_00073; -; 1.
DR  InterPro; IPR006027; Nusb.
DR  Pfam; PF01029; Nusb; 1.
KW  Transcription termination; Complete proteome.
SQ  SEQUENCE 227 AA; 25010 MW; F25A070B52342A5C CRC64;

Query Match          1.4%; Score 9; DB 1; Length 227;
Best Local Similarity 100.0%; Pred.No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  170 DSDALDSD 178
   |||||
Db  179 DSDALDSD 187

RESULT 2
RBFA_OCEIH          STANDARD;          PRT;          113 AA.
ID  RBFA_OCEIH
AC  08E0T9;
DT  10-OCT-2003 (Rel. 42, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Ribosome-binding factor A.
DE  RBFA OR OBI600.
GN  RBFA OR OBI600.

```

```

OS Oceanobacillus theysensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=HTE831 / DSM 14371 / JCM 11309;
RA MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus theysensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments";
RL Nucleic Acids Res. 30:3927-3935(2002).
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5'terminal helix region of 16S rRNA (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: Belongs to the rbfA family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP004598; BAC13556.1; -
DR HAMAP; MF_00003; -; 1.
DR InterPro; IPR000238; Rib_bind_facta.
DR Pfam; PF02033; RBFA; 1.
DR ProDom; PD007327; Rib_bind_facta; 1.
DR TIGRFAMs; TIGR00082; rbfA; 1.
DR PROSITE; PS01319; RBFA; 1.
DR Trna processing; Complete proteome.
SQ SEQUENCE 113 AA; 12924 MW; DAC02604B386F828 CRC64;

Query Match 1.3%; Score 8; DB 1; Length 113;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 606 EFDEALEQ 613
DB 93 EFDEALEQ 100

RESULT 3
FABZ_CIOAB STANDARD; PRT; 141 AA.
ID FABZ_CIOAB
AC 097DA9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ((3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase (EC 4.2.1.-)
DE ((3R)-hydroxymyristoyl ACP dehydratase).
GN FABZ OR CAC3571.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RA MEDLINE=21359325; PubMed=1146286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
CC -1- FUNCTION: Involved in saturated fatty acids biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).

```

```

CC -1- SIMILARITY: Belongs to the thioester dehydratase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE007854; AA831494.1; -
DR PIR; C9738; C97338.
DR HAMAP; MF_00406; -; 1.
DR InterPro; IPR006683; Thioestr_suff.
DR Pfam; PF03061; 4HBT; 1.
DR Lyase; Lipid synthetase; Lipid A biosynthesis; Complete proteome.
KW ACT SITE 49
FT ACT SITE 49 BY SIMILARITY.
SQ SEQUENCE 141 AA; 15438 MW; 04928B85C2B9F5A CRC64;

Query Match 1.3%; Score 7; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 213 SLSTEQI 219
DB 2 SLSTEQI 8

RESULT 4
FLAW_DESGI STANDARD; PRT; 147 AA.
ID FLAW_DESGI
AC 001056;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Flaxodxin.
DE Desulfovibrio gigas.
OS Bacteria; Proteobacteria; Delta proteobacteria; Desulfovibrionales;
OC Bacteria; Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=879;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 29494 / DSM 496;
RA MEDLINE=92329549; PubMed=1627649;
RA Helms L.R., Swenson R.P.;
RT "The primary structures of the flavodoxins from two strains of
RT Desulfovibrio gigas. Cloning and nucleotide sequence of the
RT structural genes";
RL Biochim. Biophys. Acta 1131:325-328(1992).
CC -1- FUNCTION: Low-potential electron donor to a number of redox
CC enzymes.
CC -1- COFACTOR: FMN.
CC -1- SIMILARITY: Belongs to the flavodoxin family.
CC -1- SIMILARITY: Contains 1 flavodoxin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64765; CAA46012.1; -
DR PIR; S24310; S24310.
DR HSSP; P00323; PFX2.
DR InterPro; IPR008254; Flaw_nitox_synth.
DR InterPro; IPR001094; Flavodoxin_1like.
DR InterPro; IPR001226; Flavodoxin.
DR Pfam; PF00258; flavodoxin; 1.
DR PRINTS; PR00369; FLAVODOXIN.
DR PROSITE; PS00201; FLAVODOXIN; 1.
DR PROSITE; PS50902; FLAVODOXIN_LIKE; 1.

```

KW Electron transport; Flavoprotein; FMN.  
 FT DOMAIN 4 143 FLAVODOXIN-LIKE.  
 SQ SEQUENCE 147 AA; 15186 MW; 5779A72DD139563 CRC64;  
 Query Match 1.1%; Score 7; DB 1; Length 147;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 AKVLEES 468  
 DB 23 AKVLEES 29

RESULT 5  
 COX6\_YEAST  
 ID COX6\_YEAST STANDARD; PRT; 148 AA.  
 AC P00427;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Cytochrome c oxidase polypeptide VI, mitochondrial precursor  
 DE (EC 1.9.3.1).  
 GN COX6 OR YHR051W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85080033; PubMed=6210289;  
 RA Wright R.M., Ko C., Cumsky M.G., Poyton R.O.;  
 RT Isolation and sequence of the structural gene for cytochrome c  
 RT oxidase subunit VI from *Saccharomyces cerevisiae*.  
 RL J. Biol. Chem. 259:15401-15407(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=6288c / AB972;  
 MEDLINE=94378003; PubMed=8091229;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
 Du Z., Favello A., Fulton L., Gattung S., Gelsel C., Kirsten J.,  
 Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,  
 Lucaille P., Louis E.J., Macri C., Marsden E., Meneses S., Mouser L.,  
 Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,  
 Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
 Vaudin M.;  
 RT "Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome  
 RT VIII."  
 RL Science 265:2077-2082(1994).  
 RN [3]  
 RP SEQUENCE OF 1-23 FROM N.A.  
 RX MEDLINE=89160242; PubMed=2537949;  
 RA Wright R.M., Rosenzweig B., Poyton R.O.;  
 RT "Oxidation and expression of the COX6 genetic locus in  
 RT *Saccharomyces cerevisiae*: multiple mRNAs with different 3' termini  
 RT are transcribed from COX6 and regulated differentially."  
 RL Nucleic Acids Res. 17:1103-1120(1989).  
 RN [4]  
 RP SEQUENCE OF 41-148.  
 RX MEDLINE=83030850; PubMed=6290493;  
 RA Gregor I., Tsugita A.;  
 RT "The amino acid sequence of cytochrome c oxidase subunit VI from  
 RT *Saccharomyces cerevisiae*."  
 RL J. Biol. Chem. 257:13081-13087(1982).  
 CC -1- FUNCTION: This is the heme A-containing chain of cytochrome c  
 CC oxidase, the terminal oxidase in mitochondrial electron transport.  
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome  
 CC c + 2 H(2)O.  
 CC -1- SUBUNIT: Composed of at least 11 subunits.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.  
 CC -1- SIMILARITY: Belongs to the cytochrome c oxidase Va family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M10138; AAA66900.1; -  
 DR EMBL; U00062; AAB68899.1; -  
 DR EMBL; X14452; CAA32622.1; -  
 DR PIR; A22853; OTBYC.  
 DR GenBank; 139368; -  
 DR SGD; S0001093; COX6.  
 DR GO; GO:0005751; Cytochrome chain complex IV (sensu Eukarya); IPI.  
 DR GO; GO:0004129; Cytochrome-c oxidase activity; IDA.  
 DR InterPro; IPR003204; Cyt\_c\_ox5a.  
 DR Pfam; PF02284; COX5A; 1.  
 DR Oxidoreductase; Heme; Mitochondrion; Inner membrane; Transit peptide.  
 FT TRANSIT 1 40  
 FT CHAIN 41 148  
 SQ SEQUENCE 148 AA; 17341 MW; 91F2AF286BD2FB1A CRC64;  
 Query Match 1.1%; Score 7; DB 1; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 EKEPDEA 610  
 DB 57 EKEPDEA 63

RESULT 6  
 G6PD\_SARBU  
 ID G6PD\_SARBU STANDARD; PRT; 153 AA.  
 AC Q25537;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD)  
 DE (Zwischenferment) (Fragment).  
 GN 2H.  
 OS Sarcophaga bullata (grey flesh fly) (Neobellieria bullata).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;  
 OC Sarcophagidae; Sarcophaga.  
 OX NCBI\_TaxID=7385;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Soto-Adames F.N., Robertson H.M., Berlocher S.H.;  
 RT "Phylogenetic utility of partial DNA sequences of G6PDH at different  
 RT taxonomic levels in Hexapoda with emphasis on Diptera."  
 RL Ann. Entomol. Soc. Am. 87:723-736(1994).  
 CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-  
 CC 1,5-lactone 6-phosphate + NADPH.  
 CC -1- PATHWAY: Pentose phosphate pathway; first step.  
 CC -1- SIMILARITY: Belongs to the glucose-6-phosphate dehydrogenase  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U09037; AAB02781.1; -  
 DR HSSP; P11411; IDPG.  
 DR InterPro; IPR001282; G6PD.  
 DR Pfam; PF00479; G6PD; 1.  
 DR PRINTS; PR00079; G6PDHGNASE.  
 DR ProDom; PD001129; G6PD; 1.  
 DR PROSITE; PS00069; G6P DEHYDROGENASE; PARTIAL.  
 KW Oxidoreductase; NADP; Glucose metabolism.



RT "Evolutionary studies of the nerve growth factor family reveal a  
RT novel member abundantly expressed in Xenopus ovary."  
RL Neuron 6:845-858(1991).  
CC -1- FUNCTION: Seems to promotes the survival of visceral and  
CC proprioceptive sensory neurons.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: Belongs to the NGF-beta family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M83378; AAA68880.1; -.  
DR F1R; I50400; I50400.  
DR HSSP; P20783; 1B8K.  
DR InterPro; IPR002072; NGF.  
DR Pfam; PF00243; NGF; 1.  
DR PRINTS; PRO0268; NGF; 1.  
DR ProDom; PD002052; NGF; 1.  
DR SMART; SM00140; NGF; 1.  
DR PROSITE; PS00248; NGF\_1; 1.  
DR PROSITE; PS50270; NGF\_2; 1.  
KM Growth factor; Signal\_1.  
FT SIGNAL 1 16  
FT PROPEP 17 138 POTENTIAL.  
FT CHAIN 139 257 NEUROTROPHIN-3.  
FT DISULFID 152 217 BY SIMILARITY.  
FT DISULFID 195 246 BY SIMILARITY.  
FT DISULFID 205 248 BY SIMILARITY.  
FT CARBOHYD 131 131 N-LINKED (GLCNAC..)(POTENTIAL).  
SQ SEQUENCE 257 AA; 29701 MW; EE043BA2A005C1E7 CRC64;  
Query Match 1.1%; Score 7; DB 1; Length 257;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 452 KNTLSKQ 458  
DB 45 KNTLSKQ 51  
RESULT 10  
NT3\_FELCA STANDARD; PRT; 257 AA.  
ID NT3\_FELCA  
AC 09TSY2;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Neurotrophin-3 precursor (NT-3) (Neurotrophic factor) (HDNF)  
DE (Nerve growth factor 2) (NGF-2).  
GN NTF3.  
OS Felis silvestris catus (Cat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
OX NCBI\_TaxId=9685;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20211727; PubMed=10745216;  
RA Klein E.S., Hohn A., Spatz C.J.;  
RT "Dynamic regulation of BDNF and NT-3 expression during visual system  
RT development.";  
J. Comp. Neurol. 420:1-18(2000).  
-1- FUNCTION: Seems to promotes the survival of visceral and  
CC proprioceptive sensory neurons (By similarity).  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: Belongs to the NGF-beta family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF192538; AAF03424.1; -.  
DR HSSP; P20783; 1B8K.  
DR InterPro; IPR002072; NGF.  
DR Pfam; PF00243; NGF; 1.  
DR PRINTS; PRO0268; NGF; 1.  
DR ProDom; PD002052; NGF; 1.  
DR SMART; SM00140; NGF; 1.  
DR PROSITE; PS00248; NGF\_1; 1.  
DR PROSITE; PS50270; NGF\_2; 1.  
KM Growth factor; Signal\_1.  
FT SIGNAL 1 16  
FT PROPEP 17 138 POTENTIAL.  
FT CHAIN 139 257 NEUROTROPHIN-3.  
FT DISULFID 152 217 BY SIMILARITY.  
FT DISULFID 195 246 BY SIMILARITY.  
FT DISULFID 205 248 BY SIMILARITY.  
FT CARBOHYD 131 131 N-LINKED (GLCNAC..)(POTENTIAL).  
SQ SEQUENCE 257 AA; 29403 MW; EB53F7B59C5113E4 CRC64;  
Query Match 1.1%; Score 7; DB 1; Length 257;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 452 KNTLSKQ 458  
DB 45 KNTLSKQ 51  
RESULT 11  
NT3\_HUMAN STANDARD; PRT; 257 AA.  
ID NT3\_HUMAN  
AC P20783;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Neurotrophin-3 precursor (NT-3) (Neurotrophic factor) (HDNF)  
DE (Nerve growth factor 2) (NGF-2).  
GN NTF3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90262727; PubMed=2344409;  
RA Rosenthal A., Goeddel D.V., Nguyen T., Lewis M., Shih A.,  
RA Laramie G.R., Nikolic K., Winslow J.W.;  
RT "Primary structure and biological activity of a novel human  
RT neurotrophic factor.";  
RL Neuron 4:767-773(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91045937; PubMed=2236018;  
RA Jones K.R., Reichardt L.F.;  
RT "Molecular cloning of a human gene that is a member of the nerve  
RT growth factor family.";  
J. Proc. Natl. Acad. Sci. U.S.A. 87:8060-8064(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90306351; PubMed=2365067;  
RA Kaisho Y., Yoshimura K., Nakahama K.;  
RT "Cloning and expression of a cDNA encoding a novel human neurotrophic  
RT factor.";  
FEBS Lett. 266:187-191(1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91365361; PubMed=1889806;



RA Maisonnier P.C., le Beau M.M., Espinosa R. III, Ip N.Y.,  
 RA Belluscio L., de la Monte S.M., Squinto S., Purth M.E.,  
 RA Yancopoulos G.D.;  
 RT "Human and rat brain-derived neurotrophic factor and neurotrophin-3:  
 RT gene structures, distributions, and chromosomal localizations.";  
 RL Genomics 10:558-568(1991).  
 RN [5]  
 RP SEQUENCE OF 194-236 FROM N.A.  
 RC TISSUE=Leukocyte;  
 RX MEDLINE=91222573; PubMed=2025430;  
 RA Hallboeck F., Ibanez C.F., Persson H.;  
 RT "Evolutionary studies of the nerve growth factor family reveal a  
 RT novel member abundantly expressed in Xenopus ovary.";  
 RL Neuron 6:845-858(1991).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE=95217877; PubMed=7703225;  
 RA Robinson R.C., Radziejewski C., Stuart D.I., Jones E.Y.;  
 RT "Structure of the brain-derived neurotrophic factor/neurotrophin 3  
 RT heterodimer.";  
 RL Biochemistry 34:4139-4146(1995).  
 RT [7]  
 RP VARIANT Glu-76.  
 RX MEDLINE=95251647; PubMed=7733919;  
 RA Hattori M., Nanko S.;  
 RT "Association of neurotrophin-3 gene variant with severe forms of  
 RT schizophrenia.";  
 RL Biochem. Biophys. Res. Commun. 209:513-518(1995).  
 RN [8]  
 RP VARIANT Glu-76.  
 RX MEDLINE=96253892; PubMed=8925252;  
 RA Ariami T., Takekoshi K., Itokawa M., Hamaguchi H., Toru M.;  
 RT "Failure to find associations of the CA repeat polymorphism in the  
 RT first intron and the Gly-63/Glu-63 polymorphism of the neurotrophin-3  
 RT gene with schizophrenia.";  
 RL Psychiatr. Genet. 6:13-15(1996).  
 CC -1- FUNCTION: Seems to promotes the survival of visceral and  
 CC proprioceptive sensory neurons.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Brain and peripheral tissues.  
 CC -1- POLYMORPHISM: Variant Glu-76 (frequently reported as Glu-63) was  
 CC thought to be associated with severe forms of schizophrenia. This  
 CC does not seem to be the case.  
 CC -1- SIMILARITY: Belongs to the NGF-beta family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X53655; CAA37703.1; -  
 CC EMBL; M37763; AA595953.1; -  
 CC EMBL; M61180; AA63231.1; -  
 CC PIR; A36208; C40304.  
 CC PDB; 1BND; 04-APR-96.  
 CC PDB; 1B8K; 09-FEB-99.  
 CC PDB; 1NT3; 16-JUN-00.  
 CC GeneW; HGNC:8023; NTF3.  
 CC MIM; 162660; -  
 CC GO; GO:0005102; F:receptor binding; TAS.  
 CC GO; GO:0007267; P:cell-cell signaling; TAS.  
 CC GO; GO:0007399; P:neurogenesis; TAS.  
 CC GO; GO:0007165; P:signal transduction; TAS.  
 CC InterPro; IPR002072; NGF.  
 CC Pfam; PF00243; NGF. 1.  
 CC PRINTS; PR00268; NGF.  
 CC PRODOM; PD002052; NGF. 1.  
 CC SMART; SM00140; NGF. 1.  
 CC PROSITE; PS00248; NGF\_1; 1.  
 CC PROSITE; PS50270; NGF\_2; 1.

KW Growth factor; Signal; Polymorphism; 3d-structure.  
 FT SIGNAL 1 16  
 FT PROPEP 17 138  
 FT CHAIN 17 257  
 FT DISULFID 152 217  
 FT DISULFID 152 217  
 FT DISULFID 195 246  
 FT DISULFID 205 248  
 FT CARBOHYD 131 131  
 FT VARIANT 76 76  
 FT STRAND 148 150  
 FT STRAND 154 159  
 FT STRAND 164 167  
 FT STRAND 168 169  
 FT TURN 172 175  
 FT STRAND 178 179  
 FT TURN 181 183  
 FT STRAND 186 187  
 FT STRAND 190 195  
 FT TURN 206 207  
 FT STRAND 208 208  
 FT TURN 210 211  
 FT TURN 214 230  
 FT STRAND 231 232  
 FT TURN 233 251  
 FT STRAND 257 AA; 29354 MW; 39ASB3B28E25E03 CRC64;  
 SQ SEQUENCE  
 Query Match 1.1%; Score 7; DB 1; Length 257;  
 Beat Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 452 KKKLSKQ 458  
 Db 45 KKKLSKQ 51  
 RESULT 12  
 ID NT3 MOUSE STANDARD, PRT; 258 AA.  
 AC P20181;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Neurotrophin-3 precursor (NT-3) (Neurotrophic factor) (HDNF)  
 DE (Nerve growth factor 2) (NGF-2).  
 GN NTF3 OR NTF-3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90190865; PubMed=2314473;  
 RA Hohn A., Leibrock J., Bailey K., Barde Y.-A.;  
 RT "Identification and characterization of a novel member of the nerve  
 RT growth factor/brain-derived neurotrophic factor family.";  
 RL Nature 344:339-341(1990).  
 CC -1- FUNCTION: Seems to promotes the survival of visceral and  
 CC proprioceptive sensory neurons.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Brain and peripheral tissues.  
 CC -1- SIMILARITY: Belongs to the NGF-beta family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X53257; CAA37348.1; -  
 CC PIR; S09155; S09155.

```

DR HSSP; P20783; 1B8K.
DR MGD; MGI:97380; NCF3.
DR GO; GO:000515; F:Protein binding; IMP.
DR GO; GO:0007420; P:brain development; IMP.
DR GO; GO:0008544; P:epidermal differentiation; IMP.
DR GO; GO:0007403; P:glial cell fate determination; IMP.
DR GO; GO:0007422; P:peripheral nervous system development; IMP.
DR GO; GO:0045944; P:positive regulation of transcription from P. .; IDA.
DR InterPro; IPRO02072; NGF.
DR Pfam; PF00243; NGF; 1.
DR PRINTS; PRO0268; NGF.
DR ProDom; PD002052; NGF; 1.
DR SMART; SM00140; NGF; 1.
DR PROSITE; PS00248; NGF_1; 1.
DR PROSITE; PS50270; NGF_2; 1.
KW Growth factor; Signal.
FT SIGNAL 1 16
FT PROPEP 17 139
FT CHAIN 140 258
FT DISULFID 153 218
FT DISULFID 196 247
FT DISULFID 206 249
FT CARBOHYD 131 131
SQ SEQUENCE 258 AA; 29587 MW; 7180D064E8AE6042 CRC64; (POTENTIAL).

Query Match 1.1%; Score 7; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KNTLSKQ 458
Db 45 KNTLSKQ 51

RESULT 13
NT3_RAT STANDARD; PRT; 258 AA.
AC P18280;
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Neurotrophin-3 precursor (NT-3) (Neurotrophic factor) (HDNF)
DE (Nerve growth factor 2) (NGF-2).
GN NTF3 OR NTF-3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=90319130; PubMed=2164684;
RA Ernster P., Ibanez C.F., Ebendal T., Olson L., Persson H.;
RT "Molecular cloning and neurotrophic activities of a protein with
RT structural similarities to nerve growth factor: developmental and
RT topographical expression in the brain.";
RT Proc. Natl. Acad. Sci. U.S.A. 87:5454-5458(1990).
RN [2]_
RP SEQUENCE FROM N.A.
RX MEDLINE=90208301; PubMed=2321006;
RA Maisongrosse P.C., Belluscio L., Squinto S., Ip N.Y., Furth M.E.,
RA Linday R.M., Yancopoulos G.D.;
RT "Neurotrophin-3: a neurotrophic factor related to NGF and BDNF.";
RT Science 247:1446-1451(1990).
RN [3]_
RP SEQUENCE FROM N.A.
RX MEDLINE=91365361; PubMed=1889806;
RA Maisongrosse P.C., le Beau M.M., Espinosa R. III, Ip N.Y.,
RA Belluscio L., de la Monte S.M., Squinto S., Furth M.E.,
RA Yancopoulos G.D.;
RT "Human and rat brain-derived neurotrophic factor and neurotrophin-3:
RT gene structures, distributions, and chromosomal localizations.";
RT Genomics 10:558-568(1991).
RN [4]_

```

```

RP SEQUENCE OF 195-237 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=91222573; PubMed=2025430;
RA Halboeck F., Ibanez C.F., Persson H.;
RT "Evolutionary studies of the nerve growth factor family reveal a
RT novel member abundantly expressed in Xenopus ovary.";
RL Neuron 6:845-858(1991).
CC -1- FUNCTION: Seems to promotes the survival of visceral and
CC proprioceptive sensory neurons.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Brain and peripheral tissues.
CC -1- SIMILARITY: Belongs to the NGF-beta family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL; M3643; AAA4133.1; -
DR EMBL; M3968; AAA41727.1; -
DR EMBL; M61179; AAA63497.1; -
DR PIR; A35781; A35781.
DR HSSP; P20783; 1B8K.
DR InterPro; IPRO02072; NGF.
DR Pfam; PF00243; NGF; 1.
DR PRINTS; PRO0268; NGF.
DR ProDom; PD002052; NGF; 1.
DR SMART; SM00140; NGF; 1.
DR PROSITE; PS00248; NGF_1; 1.
DR PROSITE; PS50270; NGF_2; 1.
KW Growth factor; Signal.
FT SIGNAL 1 16
FT PROPEP 17 139
FT CHAIN 140 258
FT DISULFID 153 218
FT DISULFID 196 247
FT DISULFID 206 249
FT CARBOHYD 131 131
SQ SEQUENCE 258 AA; 29644 MW; 74D557CF8518A1CE CRC64; (POTENTIAL).

Query Match 1.1%; Score 7; DB 1; Length 258;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KNTLSKQ 458
Db 45 KNTLSKQ 51

RESULT 14
CIT2_MOUSE STANDARD; PRT; 269 AA.
AC Q35740; Q35741; Q35742; Q35743; Q55198;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CBP/p300-interacting transactivator 2 (MSG-related protein 1) (MRG1
DE protein).
GN CITED2 OR MRG1 OR MSG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]_
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4).
RC STRAIN=C57BL/6 X DBA;
RX MEDLINE=98202510; PubMed=9533950;
RA Dunwoodie S.L., Rodriguez T.A., Beddington R.S.P.;
RA "Mg1 and Mrg1, founding members of a gene family, show distinct
RT patterns of gene expression during mouse embryogenesis.";

```



DR EMBL; AF109161; AAD10055.1; -.  
DR EMBL; BC004377; AA04377.1; -.  
DR Genew; HGNC:1987; CITED2.  
DR MIM; 602937; -.  
DR GO; GO:0005634; C:nucleus; NAS.  
DR GO; GO:0005515; F:protein binding; NAS.  
DR GO; GO:0003700; F:transcription factor activity; TAS.  
DR GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.  
DR InterPro; IPR007576; CITED.  
DR Pfam; PF04487; CITED; 1.  
KW Transcription regulation; Nuclear protein; Alternative splicing.  
FT DOMAIN 21 57 HIS-RICH.  
FT DOMAIN 162 199 GLY-RICH.  
FT DOMAIN 219 258 ASP/GLY-RICH (ACIDIC).  
FT VARSPLIC 159 215 Missing (in isoform 2).  
FT FTID-VSP 001089 /FTID-VSP 001089  
SQ SEQUENCE 270 AA; 28497 MW; 45DDE3A5E2B4C472 CRC64;  
Query Match 1.1%; Score 7; DB 1; Length 270;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 378 EMGLDRI 384  
Db 234 EMGLDRI 240

Search completed: April 6, 2004, 19:42:29  
Job time : 20 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:39:22 ; Search time 46 Seconds  
(without alignments)  
4286.932 Million cell updates/sec

Title: US-10-030-829-3  
Perfect score: 625  
Sequence: 1 MESSRAGPMSEKKNVQGYRP.....EFDEALEQLMYKHGLHNEDD 625

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	625	100.0	625	10 Q91DX1	Q91dx1 arabidopsis
2	13	2.1	170	10 Q7XYE6	Q7xye6 tritricum ae
3	10	1.6	304	10 Q7XY17	Q7xy17 tritricum ae
4	9	1.4	272	5 Q7Z144	Q7z144 caenorhabdi
5	9	1.4	321	16 Q8F052	Q8f052 corynebacte
6	9	1.4	382	16 Q98B79	Q98bt9 rhizobium l
7	8	1.3	212	16 Q8G635	Q8g635 bifidobacte
8	8	1.3	273	16 Q7WEP5	Q7wep5 bordetella
9	8	1.3	273	16 Q7W3C6	Q7w3c6 bordetella
10	8	1.3	273	16 Q7VU31	Q7vu31 bordetella
11	8	1.3	296	10 Q9FNU1	Q9fnu1 oryza sativ
12	8	1.3	408	17 Q8Z2M7	Q8z2m7 pyrobaculum
13	8	1.3	448	6 Q9XSE3	Q9xse3 equus caball
14	8	1.3	516	10 Q81LNS	Q81lns oryza sativ
15	8	1.3	525	16 Q8R9U3	Q8r9u3 thermoaer
16	8	1.3	586	5 Q18097	Q18097 caenorhabdi

17	8	1.3	616	5 Q96148	Q96148 plasmodium
18	8	1.3	780	5 Q811Z2	Q811z2 plasmodium
19	8	1.3	855	10 Q8S2B7	Q8s2b7 oryza sativ
20	8	1.3	916	16 Q7MG16	Q7mg16 bordetella
21	8	1.3	916	16 Q7W521	Q7w521 bordetella
22	8	1.3	916	16 Q7VVY4	Q7vv4 bordetella
23	7	1.1	69	17 Q9HPS2	Q9hps2 halobacteri
24	7	1.1	91	5 Q8T9N2	Q8t9n2 plasmodium
25	7	1.1	92	16 Q8R5Q2	Q8r5q2 thermoaer
26	7	1.1	92	16 Q8DC09	Q8dc09 vibrio vuln
27	7	1.1	93	4 Q9NS17	Q9ns17 homo sapien
28	7	1.1	94	16 Q87A19	Q87a19 xyella tas
29	7	1.1	101	17 Q9HSX0	Q9hsx0 halobacteri
30	7	1.1	108	2 Q81342	Q81342 vibrio chol
31	7	1.1	114	11 Q9D2T3	Q9d2t3 mus musculu
32	7	1.1	121	4 Q8N7K2	Q8n7k2 homo sapien
33	7	1.1	126	10 Q9MAD2	Q9mad2 arabidopsis
34	7	1.1	129	11 Q9WU16	Q9wul6 mesocricetu
35	7	1.1	135	16 Q67676	Q67676 aquilex aeo
36	7	1.1	139	6 Q8SPT8	Q8spt8 macaca mula
37	7	1.1	146	5 Q8T9Z8	Q8t9z8 plasmodium
38	7	1.1	147	13 Q7SZP9	Q7szp9 fuqua rubrip
39	7	1.1	147	16 Q8UCU6	Q8ucu6 agrobacteri
40	7	1.1	148	11 Q8BP19	Q8bp19 mus musculu
41	7	1.1	149	16 Q07614	Q07614 bacillus su
42	7	1.1	152	16 Q92T22	Q92t22 rhizobium m
43	7	1.1	157	5 Q967Z3	Q967z3 plasmodium
44	7	1.1	159	10 Q84T15	Q84t15 phaseolus a
45	7	1.1	163	5 Q7YV99	Q7yv99 trypanosoma

## ALIGNMENTS

RESULT 1	ID	Q91DX1	PRELIMINARY;	PRT;	625 AA.
AC	Q91DX1				
DT	01-OCT-2000	(TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)			
DE	EMBL CAB62356.1 (SGS3)	(Hypothetical protein).			
GN	AT5G23570.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledones; core eudicots; rosids;				
OC	euroside II; Brassicales; Brassicaceae; Arabidopsi.				
OX	NCBI_Taxid=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=COLIMBIA;				
RX	MEDLINE=20181125; PubMed=10718197;				
RA	Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,				
RA	Tabata S.;				
RT	"Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence				
RT	features of the regions of 3,076,755 bp covered by sixty Pl and TAC				
RT	clones.";				
RT	DNA Res. 7:31-63(2000).				
RL	[2]				
RN	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Columbia;				
RX	MEDLINE=20306668; PubMed=10850495;				
RA	Mourrain P., Beclin C., Elmayan T., Feuerbach F., Godon C.,				
RA	Morel U.-B., Jouette D., Lacombe A.-M., Nikic S., Picault N.,				
RA	Remoue K., Sanjal M., Vo T.-A., Vaucheret H.;				
RT	"Arabidopsis SGS2 and SGS3 Genes are Required for Posttranscriptional				
RT	Gene Silencing and Natural Virus Resistance.";				
RL	Cell 101:533-542(2000).				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RA	Yamada K., Chan M.-M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,				
RA	Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,				
RA	Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,				

RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shimizu P.,  
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.,  
 RT "Arabidopsis Full Length cDNA Clones";  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Heuan V.W., Lee J.M.,  
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,  
 RA Yu G., Yuan S., Cammici P., Chen H., Cheuk R., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shimizu P.,  
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.,  
 RT "Arabidopsis Open Reading Frame (ORF) Clones";  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB025633; BAA97244.1; -;  
 DR EMBL; AF239719; AAF73960.1; -;  
 DR EMBL; BT002944; AAO22757.1; -;  
 DR EMBL; BT004380; AAO42374.1; -;  
 DR InterPro; IPR005380; XS.  
 DR InterPro; IPR005381; zF.  
 DR Pfam; PF03468; XS; 1.  
 DR Pfam; PF03470; zF-XS; 1.  
 DR Hypochemical Protein;  
 KW SEQUENCE 625 AA; 71971 MW; 456E2A1396706A96 CRC64;  
 SQ

Query Match 100.0%; Score 625; DB 10; Length 625;  
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 625; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MSSRAGPMSKKNVGGYRPEVEQLVQGLAGTRLASDGDGGEVITSKKNKPKNTSG 60  
 1 MSSRAGPMSKKNVGGYRPEVEQLVQGLAGTRLASDGDGGEVITSKKNKPKNTSG 60  
 61 KTWVSONSNPPRANGGQOQGRGNSVSGRNNVSGRNGNGGICQANTSGRALSRYKDN 120  
 61 KTWVSONSNPPRANGGQOQGRGNSVSGRNNVSGRNGNGGICQANTSGRALSRYKDN 120  
 61 KTWVSONSNPPRANGGQOQGRGNSVSGRNNVSGRNGNGGICQANTSGRALSRYKDN 120  
 121 NFVAPPVSRPPELEGGMNQAAGSGAHTAVQEPFVDEDDVDNASEENDSDALDDSD 180  
 121 NFVAPPVSRPPELEGGMNQAAGSGAHTAVQEPFVDEDDVDNASEENDSDALDDSD 180  
 121 NFVAPPVSRPPELEGGMNQAAGSGAHTAVQEPFVDEDDVDNASEENDSDALDDSD 180  
 181 IASDDVDSDVQSHSGSRKONKFKFPGSLDLSLSEIQINERQRMCPACQNGGALDW 240  
 181 IASDDVDSDVQSHSGSRKONKFKFPGSLDLSLSEIQINERQRMCPACQNGGALDW 240  
 181 IASDDVDSDVQSHSGSRKONKFKFPGSLDLSLSEIQINERQRMCPACQNGGALDW 240  
 241 YNLHPLAARTGARRVTLHRELAEVLKEDLOMRGASVPCGEITYGQMKLGDEKDY 300  
 241 YNLHPLAARTGARRVTLHRELAEVLKEDLOMRGASVPCGEITYGQMKLGDEKDY 300  
 241 YNLHPLAARTGARRVTLHRELAEVLKEDLOMRGASVPCGEITYGQMKLGDEKDY 300  
 301 IVPMPWIIINNTRLDKDNDKWLGMGNQELLEFDYKELRAPHSTGPGCHRGMSVME 360  
 301 IVPMPWIIINNTRLDKDNDKWLGMGNQELLEFDYKELRAPHSTGPGCHRGMSVME 360  
 301 IVPMPWIIINNTRLDKDNDKWLGMGNQELLEFDYKELRAPHSTGPGCHRGMSVME 360  
 361 SSATGYEABRLHRELAEVLKEDLOMRGASVPCGEITYGQMKLGDEKDY 420  
 361 SSATGYEABRLHRELAEVLKEDLOMRGASVPCGEITYGQMKLGDEKDY 420  
 361 SSATGYEABRLHRELAEVLKEDLOMRGASVPCGEITYGQMKLGDEKDY 420  
 421 TRLFELKSYQEMVVKELRQISEDNQOLNTFKNLSKONKHAVALSESLSEKLRRA 480  
 421 TRLFELKSYQEMVVKELRQISEDNQOLNTFKNLSKONKHAVALSESLSEKLRRA 480  
 421 TRLFELKSYQEMVVKELRQISEDNQOLNTFKNLSKONKHAVALSESLSEKLRRA 480  
 481 EDNRIVRQTYQEHQNEENDADRPFNDSIKQIHRRDPAKENFEMLOOQERAVYQ 540  
 481 EDNRIVRQTYQEHQNEENDADRPFNDSIKQIHRRDPAKENFEMLOOQERAVYQ 540  
 481 EDNRIVRQTYQEHQNEENDADRPFNDSIKQIHRRDPAKENFEMLOOQERAVYQ 540  
 541 OQONINPSSNDCCRAAEVSSFLFOEKEMEEFEVEREMLIKQOEKEMDMKKRHEE 600  
 541 OQONINPSSNDCCRAAEVSSFLFOEKEMEEFEVEREMLIKQOEKEMDMKKRHEE 600  
 541 OQONINPSSNDCCRAAEVSSFLFOEKEMEEFEVEREMLIKQOEKEMDMKKRHEE 600  
 601 FDLKEFDEALQMLYKHLNEDD 625

DB 601 FDLKEFDEALQMLYKHLNEDD 625

RESULT 2  
 ID 07XYE6 PRELIMINARY; PRT; 170 AA.  
 AC 07XYE6;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE HOIR (Fragment).  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticaceae; Triticum.  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. PH 82-2-2;  
 RA Li J.R., Wang F., Li O.Z., Zhang X.S.;  
 RT "Gene Isolation and expression of a new Zn-finger";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF469493; AAP80610.1; -;  
 FT NON TER 1  
 SQ SEQUENCE 170 AA; 19980 MW; 0AFD21F1778B3068 CRC64;

Query Match 2.1%; Score 13; DB 10; Length 170;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

322 WLMGNQELLEVF 334  
 1 WLMGNQELLEVF 13

RESULT 3  
 ID 07XY17 PRELIMINARY; PRT; 304 AA.  
 AC 07XY17;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Emr1.  
 OS Triticum aestivum (Wheat).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OX NCBI\_TaxID=4565;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. PH 82-2-2;  
 RA Zhao X., Li O., Zhang X.;  
 RT "Isolation and expression of a new kind of gene involve in  
 embryogenesis in Triticum aestivum L.";  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF542974; AAP8062.1; -;  
 SQ SEQUENCE 304 AA; 35958 MW; 89980215A9584208 CRC64;

Query Match 1.6%; Score 10; DB 10; Length 304;  
 Best Local Similarity 100.0%; Pred. No. 0.15;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

325 MGNQELLEVF 334  
 1 MGNQELLEVF 10

RESULT 4  
 ID 07Z144 PRELIMINARY; PRT; 272 AA.  
 AC 07Z144;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)

DT 01-OCT-2003 (TREMBlrel. 25, last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, last annotation update)  
 DE Hypothetical protein ZK546.17.  
 GN ZK546.17.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA Wilson R.;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX Hallsworth K.;  
 RT "The sequence of C. elegans cosmid ZK546";  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX Waterston R.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX Wilson R.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U9380; AAP68926.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 272 AA; 31290 MW; CB99BA62BD7DEB84 CRC64;  
 QY  
 DB 469 LEIMSEKLR 477  
 DB 102 LEIMSEKLR 110  
 RESULT 5  
 Q8FO52 PRELIMINARY; PRT; 321 AA.  
 AC Q8FO52;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)  
 DE Putative glucose dehydrogenase-B.  
 GN CE1282.  
 OS Corynebacterium efficiens.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.  
 OC NCBI\_TaxID=152794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
 RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,  
 RA Ikeno K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,  
 RA Ueda Y., Sugimoto S.;  
 RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AP005218; BAC18092.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 321 AA; 34379 MW; C3D9C980AFF3C080 CRC64;  
 QY  
 DB 1.4%; Score 9; DB 16; Length 321;  
 Best Local Similarity 100.0%; Pred. No. 1.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 TAEDNRIVR 487  
 DB 75 TAEDNRIVR 83  
 RESULT 6  
 Q98BT9 PRELIMINARY; PRT; 382 AA.  
 AC Q98BT9;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, last annotation update)  
 DE Acyl-CoA dehydrogenase.  
 GN MLI5430.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OC NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Matsumoto M., Iida S., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP003006; BAB51883.1; -  
 DR GO: GO:0003995; F:acyl-CoA dehydrogenase activity; IEA.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0006118; P:electron transport; IEA.  
 DR InterPro: IPR006090; Acyl-CoA\_dh\_C.  
 DR InterPro: IPR006091; Acyl-CoA\_dh\_M.  
 DR Pfam: PF02770; Acyl-CoA\_dh\_M; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 382 AA; 41252 MW; DB2B4A400E1864C7 CRC64;  
 QY  
 DB 261 HRELAIVLE 269  
 DB 16 HRELAIVLE 24  
 RESULT 7  
 Q8G635 PRELIMINARY; PRT; 212 AA.  
 AC Q8G635;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)  
 DE Hypothetical protein.  
 GN BL0815.  
 OS Bifidobacterium longum.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;  
 OC Bifidobacteriaceae; Bifidobacterium.  
 OC NCBI\_TaxID=216816;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCC 2705;  
 RX MEDLINE=22294977; PubMed=12381787;  
 RA Schell M.A., Karmaliantzou M., Snel B., Vilianova D., Berger B.,  
 RA Pessi G., Zehren M.-C., Desiere F., Bork P., Delley M.,  
 RA Pridmore R.D., Arigoni F.;  
 RT "The genome sequence of Bifidobacterium longum reflects its adaptation  
 RT to the human gastrointestinal tract.";



```

RL  Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).
DR  EMBL; AF014703; AM24629.1; -.
KW  Hypothetical protein. Complete proteome.
SQ  SEQUENCE 212 AA; 25137 MW; EDA495CD83DDB53D CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 212;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 ASDDYDSD 189
DB 195 ASDDYDSD 202

RESULT 8
Q7WEP5 PRELIMINARY; PRT; 273 AA.
ID Q7WEP5
AC Q7WEP5;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative enoyl-CoA hydratase.
GN B4589.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxId=518;

SEQUENCE FROM N.A.
RP STRAIN=RB50 / ATCC BAA-589;
RC MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Basher D., Cherevach I.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Skelton J., Squares R., Squares S., Stevens K.,
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40 (2003).
DR EMBL; BX640451; CAE34951.1; -.
KW Complete proteome.
SQ SEQUENCE 273 AA; 29859 MW; EA0423C75C3AC225 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 273;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 ERLHRELA 377
DB 45 ERLHRELA 52

RESULT 9
Q7W3C6 PRELIMINARY; PRT; 273 AA.
ID Q7W3C6
AC Q7W3C6;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative enoyl-CoA hydratase.
GN BPA119.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxId=519;
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;

```

```

RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Basher D., Cherevach I.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Skelton J., Squares R., Squares S., Stevens K.,
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40 (2003).
DR EMBL; BX640435; CAE39398.1; -.
KW Complete proteome.
SQ SEQUENCE 273 AA; 29902 MW; 692511C67C253C3A CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 273;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 ERLHRELA 377
DB 45 ERLHRELA 52

RESULT 10
Q7VU31 PRELIMINARY; PRT; 273 AA.
ID Q7VU31
AC Q7VU31;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative enoyl-CoA hydratase.
GN BP3304.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxId=520;

SEQUENCE FROM N.A.
RP STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RC MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Basher D., Cherevach I.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Skelton J., Squares R., Squares S., Stevens K.,
RA "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40 (2003).
DR EMBL; BX640421; CAE43569.1; -.
KW Complete proteome.
SQ SEQUENCE 273 AA; 29859 MW; EA0423C75C3AC225 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 273;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 370 ERLHRELA 377
DB 45 ERLHRELA 52

RESULT 11
Q9FN11 PRELIMINARY; PRT; 296 AA.
ID Q9FN11

```

```

AC Q9FNUL;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE 3615.5.
OS Oryza sativa (rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare; TRANSPOSON=gaigan;
RX MEDLINE=21140318; PubMed=11244114;
RA Dubcovsky J., Ramakrishna W., SanMiguel P.J., Bussio C.S., Yan L.,
RA Shiloff B.A., Bennett J.L.
RT "Comparative genome analysis of colinear barley and rice bacterial
RT artificial chromosomes."
RL Plant Physiol. 125:1342-1353(2001).
DR EMBL; AY013245; AAC45493.1; -.
DR Gramene; Q9FNUL; -.
SQ SEQUENCE 296 AA; 3326 MW; CC472BACED677608 CRC64;

Query Match 1.3%; Score 8; DB 10; Length 296;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 SDALDSD 178
Db 254 SDALDSD 261

RESULT 12
Q8Z2M7 PRELIMINARY; PRT; 408 AA.
ID Q8Z2M7
AC Q8Z2M7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein PAE0172.
GN PAE0172.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Filiz-Gibson S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum."
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AE009754; AAL62612.1; -.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR SMART; SM00382; AAA; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 408 AA; 46291 MW; 02A24E5A2523386F CRC64;

Query Match 1.3%; Score 8; DB 17; Length 408;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 261 HRELAEVL 268
Db 86 HRELAEVL 93

RESULT 13
Q9XS63 PRELIMINARY; PRT; 448 AA.
ID Q9XS63

```

```

AC Q9XS63;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Chromogranin A.
GN CGA.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal medulla;
RX MEDLINE=20492714; PubMed=11039590;
RA Sato F., Hasegawa T., Katayama Y., Iwanaga T., Yanaihara N., Kanno T.,
RA Ichida N.
RT "Equus caballus CGA mRNA for chromogranin A, complete cds."
RL J. Vet. Med. Sci. 62:953-959(2000).
DR EMBL; AB025570; BAA76748.1; -.
DR InterPro; IPR001819; Chromogranin_AB.
DR InterPro; IPR001990; Granin.
DR Pfam; PF01271; Granin; 1.
DR PRINTS; PR00659; CHROMOGRANIN.
DR PROSITE; PS00422; GRANINS_1; 1.
DR PROSITE; PS00423; GRANINS_2; 1.
SQ SEQUENCE 448 AA; 49861 MW; EC2D6418F5BA5274 CRC64;

Query Match 1.3%; Score 8; DB 6; Length 448;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 ELAEVLEK 270
Db 102 ELAEVLEK 109

RESULT 14
Q8L1N5 PRELIMINARY; PRT; 516 AA.
ID Q8L1N5
AC Q8L1N5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein OJ112 G08.17.
GN OSJNBA0032E21.01 OR OJ1112 G08.17.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_TaxID=3947;
RN [1]
RP SEQUENCE FROM N.A.
RA Eastman A.P., Smith S.C., Bertin N., Liang C., Najjar F.Z., Pratt L.H.,
RA Cordouanier-Pratt M.-M.;
RT "Uncloned."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RX Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA Overton II L.L., Talcott T., Kim M.M., Bera J.J., Jin S.S.,
RA Padrosh D.W., Tallon L.J., Koo H., Ziemann V., Heiao J., Blunt S.,
RA Venaken S.S., Riedmuller S.B., Uterback T.T., Feldlyum T.V.,
RA Yang Q.Q., Haas B.J., Sun B.B., Peterson J.D., Quackenbush J.J.,
RA White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OJ112 G08 genomic sequence."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RX Buell R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF377947; AAM34395.2; -.

```

DR EMBL: AC135225; AAP68355.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 516 AA; 58268 MW; 0B9B6A7D0273F55 CRC64;

Query Match 1.3%; Score 8; DB 10; Length 516;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 SDALDDSD 178  
 |||||  
 DB 307 SDALDDSD 314

## RESULT 15

Q8R9U3 PRELIMINARY; PRT; 525 AA.  
 AC Q8R9U3;  
 DT 01-JUN-2002 (Tremblere1.21, Created)  
 DT 01-JUN-2002 (Tremblere1.21, Last sequence update)  
 DT 01-JUN-2003 (Tremblere1.24, Last annotation update)  
 DE Predicted kinase related to dihydroxyacetone kinase.  
 GN TTE1493.  
 OS Thermomanaerobacter tengcongensis.  
 OC Bacteria; Firmicutes; Clostridia; Thermomanaerobacteriales;  
 OC Thermomanaerobacteriaceae; Thermomanaerobacter.  
 OC NCBI\_TaxID=119072;  
 OX  
 RN  
 RP  
 RC STRAIN=MB4 / JCM 11007;  
 RX MEDLINE=21992816; PubMed=11997336;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 Tan H., Chen R., Wang J., Yu J., Yang H.,  
 "A complete sequence of T. tengcongensis genome."  
 RT Genome Res. 12:689-700(2002).  
 RL EMBL: AE013106; AM24711.1; -.  
 DR GO: 0004371; P:glycerol kinase activity; IEA.  
 DR GO: 0006071; P:glycerol metabolism; IEA.  
 DR InterPro: IPR004007; Dsk2.  
 DR Pfam: PF02734; Dsk2; 1.  
 KW Kinase; Complete proteome.  
 SQ SEQUENCE 525 AA; 58638 MW; 6FB4C8215FFDEB2C CRC64;

Query Match 1.3%; Score 8; DB 16; Length 525;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 602 DLEKEPDE 609  
 |||||  
 DB 405 DLEKEPDE 412

Search completed: April 6, 2004, 19:43:28  
 Job time : 48 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 19:40:52 ; Search time 23 Seconds  
(without alignments)  
1402.880 Million cell updates/sec

Title: US-10-030-829-3

Perfect score: 625  
Sequence: 1 MESSRAGPMSEKKNVQGYRP.....EFDEALQLMYKHGLNEDD 625

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- 1: Issued Patents\_AA:\*
- 2: /cgm2\_6/prodata/2/1aa/5A\_COMB.pep:\*
- 3: /cgm2\_6/prodata/2/1aa/5B\_COMB.pep:\*
- 4: /cgm2\_6/prodata/2/1aa/6A\_COMB.pep:\*
- 5: /cgm2\_6/prodata/2/1aa/6B\_COMB.pep:\*
- 6: /cgm2\_6/prodata/2/1aa/6C\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.3	327	4 US-09-252-991A-30948	Sequence 30948, A
2	8	1.3	418	4 US-09-489-039A-8728	Sequence 8728, Ap
3	8	1.3	521	4 US-09-543-681A-5969	Sequence 5969, Ap
4	7	1.1	85	4 US-09-543-681A-4934	Sequence 4934, Ap
5	7	1.1	94	2 US-08-629-291A-31	Sequence 31, Appl
6	7	1.1	94	2 US-08-658-335B-31	Sequence 31, Appl
7	7	1.1	94	4 US-09-406-640-31	Sequence 31, Appl
8	7	1.1	102	4 US-09-543-681A-7823	Sequence 7823, Ap
9	7	1.1	108	4 US-09-252-991A-28062	Sequence 28062, A
10	7	1.1	162	4 US-09-252-991A-22093	Sequence 22093, A
11	7	1.1	175	4 US-09-705-621-13	Sequence 13, Appl
12	7	1.1	233	4 US-09-543-681A-4354	Sequence 4354, Ap
13	7	1.1	257	1 US-08-451-947-4	Sequence 4, Appl1
14	7	1.1	257	2 US-08-424-826A-4	Sequence 4, Appl1
15	7	1.1	257	3 US-08-910-691-7	Sequence 7, Appl1
16	7	1.1	257	3 US-08-928-694-4	Sequence 4, Appl1
17	7	1.1	257	4 US-08-450-842-4	Sequence 4, Appl1
18	7	1.1	257	4 US-08-451-390-4	Sequence 4, Appl1
19	7	1.1	257	5 PCT-US91-06950-4	Sequence 4, Appl1
20	7	1.1	283	4 US-09-434-774-6	Sequence 6, Appl1
21	7	1.1	288	4 US-09-813-918-3	Sequence 3, Appl1
22	7	1.1	295	4 US-09-134-000C-5063	Sequence 5063, Ap
23	7	1.1	300	4 US-09-489-039A-12934	Sequence 12934, A
24	7	1.1	312	4 US-09-252-991A-31204	Sequence 31204, A
25	7	1.1	351	3 US-08-688-988-28	Sequence 28, Appl
26	7	1.1	361	4 US-09-489-039A-12349	Sequence 12349, A
27	7	1.1	375	4 US-09-107-532A-4312	Sequence 4312, Ap

28	7	1.1	389	4 US-09-489-039A-10992	Sequence 10992, A
29	7	1.1	395	4 US-09-673-395A-612	Sequence 612, App
30	7	1.1	445	4 US-09-252-991A-28986	Sequence 28986, A
31	7	1.1	453	4 US-09-711-164-349	Sequence 349, App
32	7	1.1	454	4 US-09-813-918-2	Sequence 2, Appl1
33	7	1.1	456	4 US-09-543-681A-5516	Sequence 5516, Ap
34	7	1.1	511	4 US-09-252-991A-28608	Sequence 28608, A
35	7	1.1	519	4 US-09-489-039A-11591	Sequence 11591, A
36	7	1.1	528	4 US-09-356-806-8	Sequence 8, Appl1
37	7	1.1	530	3 US-09-180-852-2	Sequence 2, Appl1
38	7	1.1	530	4 US-09-356-806-113	Sequence 113, App
39	7	1.1	532	4 US-09-252-991A-29652	Sequence 29652, A
40	7	1.1	577	2 US-08-756-317-13	Sequence 13, Appl1
41	7	1.1	578	3 US-08-981-215-1	Sequence 1, Appl1
42	7	1.1	589	4 US-09-543-681A-6155	Sequence 6155, Ap
43	7	1.1	596	4 US-09-252-991A-26463	Sequence 26463, A
44	7	1.1	630	4 US-09-252-991A-19822	Sequence 19822, A
45	7	1.1	633	4 US-09-252-991A-26576	Sequence 26576, A

# ALIGNMENTS

```

RESULT 1
US-09-252-991A-30948
Sequence 30948, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 30948
LENGTH: 327
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30948

Query Match      1.3%; Score 8; DB 4; Length 327;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      109 GGRALSR 116
DB      113 GGRALSR 120

RESULT 2
US-09-489-039A-8728
Sequence 8728, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709 2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 8728
LENGTH: 418
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8728

```

Query Match 1.3%; Score 8; DB 4; Length 418;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 GARRVRLH 261  
 DB 373 GARRVRLH 380

RESULT 3  
 US-09-543-681A-5969  
 ; Sequence 5969, Application US/09543681A

; Patent No. 6605709  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GARY BRETON  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
 ; FILE REFERENCE: 2709, 1002-001  
 ; CURRENT APPLICATION NUMBER: US/09/543,681A  
 ; CURRENT FILING DATE: 2000-04-05  
 ; PRIOR APPLICATION NUMBER: US 60/128,706  
 ; PRIOR FILING DATE: 1999-04-09  
 ; NUMBER OF SEQ ID NOS: 8344  
 ; SEQ ID NO 5969  
 ; LENGTH: 521  
 ; TYPE: PRT  
 ; ORGANISM: Proteus mirabilis  
 ; US-09-543-681A-5969

Query Match 1.3%; Score 8; DB 4; Length 521;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GGWEVTS 48  
 DB 494 GGWEVTS 501

RESULT 4  
 US-09-543-681A-4934  
 ; Sequence 4934, Application US/09543681A

; Patent No. 6605709  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GARY BRETON  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
 ; FILE REFERENCE: 2709, 1002-001  
 ; CURRENT APPLICATION NUMBER: US/09/543,681A  
 ; CURRENT FILING DATE: 2000-04-05  
 ; PRIOR APPLICATION NUMBER: US 60/128,706  
 ; PRIOR FILING DATE: 1999-04-09  
 ; NUMBER OF SEQ ID NOS: 8344  
 ; SEQ ID NO 4934  
 ; LENGTH: 85  
 ; TYPE: PRT  
 ; ORGANISM: Proteus mirabilis  
 ; US-09-543-681A-4934

Query Match 1.1%; Score 7; DB 4; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 EINSEKL 476  
 DB 13 EINSEKL 19

RESULT 5  
 US-08-629-291A-31  
 ; Sequence 31, Application US/08629291A  
 ; Patent No. 5959174  
 ; GENERAL INFORMATION:

; APPLICANT: Coruzzi, Gloria  
 ; APPLICANT: Oliveira, Igor  
 ; APPLICANT: Lam, Hon-Ming  
 ; APPLICANT: Hsieh, Ming-Hsiun  
 ; TITLE OF INVENTION: PLANT GLUTAMATE RECEPTORS  
 ; NUMBER OF SEQUENCES: 36  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/629,291A  
 ; FILING DATE: 08-APR-1996  
 ; CLASSIFICATION: 800  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 5914-050  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 869-9740/8864  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 31:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 94 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-629-291A-31

Query Match 1.1%; Score 7; DB 2; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 ONINPSS 549  
 DB 64 ONINPSS 70

RESULT 6  
 US-08-658-335B-31  
 ; Sequence 31, Application US/08658335B  
 ; Patent No. 5981703  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Coruzzi, Gloria  
 ; APPLICANT: Oliveira, Igor  
 ; APPLICANT: Lam, Hon-Ming  
 ; APPLICANT: Hsieh, Ming-Hsiun  
 ; TITLE OF INVENTION: PLANT GLUTAMATE RECEPTORS  
 ; NUMBER OF SEQUENCES: 36  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/658,335B

FILING DATE: 05-JUN-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 5914-052  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 94 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-658-335B-31

Query Match 1.1%; Score 7; DB 2; Length 94;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 QINPSS 549  
DB 64 QINPSS 70

RESULT 7  
US-09-406-640-31  
Sequence 31, Application US/09406640  
Patent No. 6451546

GENERAL INFORMATION:  
APPLICANT: Coruzzi, Gloria

Oliveira, Igor  
Lam, Hon-Ming  
Hsieh, Ming-Hsiun

TITLE OF INVENTION: PLANT GLUTAMATE RECEPTORS  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas

CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/406,640  
FILING DATE: 27-Sep-1999

CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 5914-082  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:  
LENGTH: 94 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 31:

US-09-406-640-31

Query Match 1.1%; Score 7; DB 4; Length 94;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 QINPSS 549  
DB 64 QINPSS 70

RESULT 8  
US-09-543-681A-7823

Sequence 7823, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:

APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
FILE REFERENCE: 2709,1002-001

CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05

PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09

NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 7823  
LENGTH: 102

TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-7823

Query Match 1.1%; Score 7; DB 4; Length 102;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 KJHRELA 265  
DB 59 KJHRELA 65

RESULT 9  
US-09-252-991A-28062

Sequence 28062, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196,136

CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 28062  
LENGTH: 108

TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28062

Query Match 1.1%; Score 7; DB 4; Length 108;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 ARGSAQ 147  
DB 42 ARGSAQ 48

RESULT 10  
US-09-252-991A-22093  
Sequence 22093, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 22093  
LENGTH: 162  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22093

Query Match 1.1%; Score 7; DB 4; Length 162;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 537 VVGQOQO 543  
DB 141 VVGQOQO 147

RESULT 11  
US-09-705-621-13  
Sequence 13, Application US/09705621  
Patent No. 667335  
GENERAL INFORMATION:  
APPLICANT: ESTES, MARY  
TITLE OF INVENTION: ROTAVIRUS ENTEROTOXIN NSP4 AND METHODS OF USING SAME  
FILE REFERENCE: P01932US3  
CURRENT APPLICATION NUMBER: US/09/705,621  
CURRENT FILING DATE: 2000-11-03  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: Patent version 3.0  
SEQ ID NO 13  
LENGTH: 175  
TYPE: PRT  
ORGANISM: Rotavirus  
US-09-705-621-13

Query Match 1.1%; Score 7; DB 4; Length 175;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 434 VVKELRQ 440  
DB 102 VVKELRQ 108

RESULT 12  
US-09-543-681A-4354  
Sequence 4354, Application US/09543681A  
Patent No. 6605709  
GENERAL INFORMATION:  
APPLICANT: GARY BRETON  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.1002-001  
CURRENT APPLICATION NUMBER: US/09/543,681A  
CURRENT FILING DATE: 2000-04-05  
PRIOR APPLICATION NUMBER: US 60/128,706  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 8344  
SEQ ID NO 4354  
LENGTH: 233  
TYPE: PRT  
ORGANISM: Proteus mirabilis  
US-09-543-681A-4354

Query Match 1.1%; Score 7; DB 4; Length 233;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 KLSKQNK 460  
DB 162 KLSKQNK 168

RESULT 13  
US-08-451-947-4  
Sequence 4, Application US/08451947  
Patent No. 5702906  
GENERAL INFORMATION:  
APPLICANT: GENENTECH, INC.  
TITLE OF INVENTION: NOVEL NEUROTROPHIC FACTOR  
NUMBER OF SEQUENCES: 100  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: palin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/451,947  
FILING DATE:

CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/426419  
FILING DATE: 19-APR-1995  
PRIOR APPLICATION DATA: 08/030013  
APPLICATION NUMBER: 22-MAR-1993  
PRIOR APPLICATION DATA: 07/648482  
APPLICATION NUMBER: 31-JAN  
FILING DATE: 07/587707  
APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:  
NAME: Torchia, Timothy E.  
REGISTRATION NUMBER: 36,700  
REFERENCE/DOCKET NUMBER: 666P2C1D2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/952-8674  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 257 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-451-947-4

Query Match 1.1%; Score 7; DB 1; Length 257;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KXKLSQ 458  
DB 45 KXKLSQ 51

RESULT 14  
US-08-424-826A-4

; Sequence 4, Application US/08424826A  
; Patent No. 5830858  
; GENERAL INFORMATION:  
; APPLICANT: Rosenthal, Arnon  
; TITLE OF INVENTION: NOVEL NEUTROPHILIC FACTOR  
; NUMBER OF SEQUENCES: 98  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WinPatIn (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/424,826A  
; FILING DATE: 19-Apr-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/240387  
; FILING DATE: 10-May-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/648482  
; FILING DATE: 31-Jan-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/587707  
; FILING DATE: 25-SEP-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Torchia, PhD., Timothy E.  
; REGISTRATION NUMBER: 36,700  
; REFERENCE/DOCKET NUMBER: P0666P1C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-8674  
; TELEFAX: 910/371-7168  
; TELEEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 257 amino acids  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
; US-08-424-826A-4

Query Match 1.1%; Score 7; DB 2; Length 257;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KKKLSKQ 458  
DB 45 KKKLSKQ 51

RESULT 15  
US-08-910-691-7  
; Sequence 7, Application US/08910691  
; Patent No. 6015552  
; GENERAL INFORMATION:  
; APPLICANT: WATANABE, Tatsuya  
; APPLICANT: YOSHITOMI, Sumie  
; APPLICANT: SASADA, Reiko  
; TITLE OF INVENTION: THERAPEUTIC AGENT FOR NEUTROPENIA  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02109

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,691  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/074,969  
; FILING DATE: 19930604  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NEUNER, George W  
; REGISTRATION NUMBER: 26964  
; REFERENCE/DOCKET NUMBER: 12345  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)523-3400  
; TELEFAX: (617)523-6440  
; TELEEX: 200291 STRE UR  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 257 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-910-691-7

Query Match 1.1%; Score 7; DB 3; Length 257;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 KKKLSKQ 458  
DB 45 KKKLSKQ 51

Search completed: April 6, 2004, 19:44:36  
Job time : 24 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: April 6, 2004, 19:43:32 ; Search time 46 Seconds  
(without alignments)  
3568.388 Million cell updates/sec

Title: US-10-030-829-3  
Perfect score: 625  
Sequence: 1 MSRRAGPMSEKKNVQGGYRP.....EPDEALEQMLYKHLNEDD 625

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1071772 seqs, 26263353 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

## Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PC7\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PC7US\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	1.4	223	12	US-10-424-599-234019
2	9	1.4	227	9	US-09-738-626-5288
3	9	1.4	324	12	US-10-424-599-234018
4	9	1.4	382	15	US-10-369-493-12141
5	9	1.4	611	12	US-10-425-114-58390
6	8	1.3	916	12	US-10-282-122A-51413
7	7	1.1	18	14	US-10-225-567A-1514
8	7	1.1	36	9	US-09-205-658-225
9	7	1.1	36	10	US-09-963-693-225
10	7	1.1	56	12	US-10-424-599-250491
11	7	1.1	56	14	US-10-029-386-31679
12	7	1.1	57	15	US-10-264-049-3845
13	7	1.1	60	12	US-10-424-599-154923
14	7	1.1	62	12	US-10-424-599-228124
15	7	1.1	68	12	US-10-424-599-169910

16	7	1.1	69	14	US-10-106-698-5320	Sequence 5320, Ap
17	7	1.1	78	12	US-10-424-599-280255	Sequence 280255,
18	7	1.1	94	14	US-10-223-047-31	Sequence 31, Appl
19	7	1.1	104	9	US-09-764-869-650	Sequence 650, App
20	7	1.1	104	15	US-10-091-504-650	Sequence 650, App
21	7	1.1	104	15	US-10-227-577-650	Sequence 650, App
22	7	1.1	121	15	US-10-108-260A-4862	Sequence 4862, Ap
23	7	1.1	136	12	US-10-425-114-56375	Sequence 56375, A
24	7	1.1	136	12	US-10-425-114-69428	Sequence 69428, A
25	7	1.1	138	12	US-10-424-599-177015	Sequence 177015,
26	7	1.1	138	14	US-10-155-886-53	Sequence 53, Appl
27	7	1.1	138	14	US-10-155-886-59	Sequence 59, Appl
28	7	1.1	139	14	US-10-155-886-55	Sequence 55, Appl
29	7	1.1	139	14	US-10-155-886-57	Sequence 57, Appl
30	7	1.1	141	12	US-10-282-122A-51842	Sequence 51842, A
31	7	1.1	148	15	US-10-369-493-22038	Sequence 22038, A
32	7	1.1	166	12	US-10-425-114-72748	Sequence 72748, A
33	7	1.1	173	12	US-10-424-599-277920	Sequence 277920,
34	7	1.1	180	12	US-10-425-114-54097	Sequence 54097, A
35	7	1.1	189	12	US-10-424-599-283874	Sequence 283874,
36	7	1.1	196	12	US-10-425-114-64832	Sequence 64832, A
37	7	1.1	211	12	US-10-627-476-128	Sequence 128, App
38	7	1.1	216	9	US-09-861-451A-26	Sequence 26, Appl
39	7	1.1	219	12	US-10-282-122A-69066	Sequence 69066, A
40	7	1.1	226	12	US-10-335-977-5666	Sequence 5666, Ap
41	7	1.1	227	12	US-10-335-977-5665	Sequence 5665, Ap
42	7	1.1	253	14	US-10-029-386-33003	Sequence 33003, A
43	7	1.1	257	8	US-08-450-842-4	Sequence 4, Appl
44	7	1.1	257	10	US-09-788-188-5	Sequence 5, Appl
45	7	1.1	257	10	US-09-788-188-6	Sequence 6, Appl

## ALIGNMENTS

RESULT 1  
US-10-424-599-234019  
; Sequence 234019, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424, 599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 234019  
; LENGTH: 223  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(223)  
; OTHER INFORMATION: unsure at all Xaa locations  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_53343C.1.pep  
; US-10-424-599-234019

Query Match 1.4%; Score 9; DB 12; Length 223;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 ROWHCPACQ 232  
Db 119 ROWHCPACQ 127

RESULT 2  
US-09-738-626-5288

```

; Sequence 5288, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIYOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5288
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5288

```

```

Query Match      1.4%; Score 9; DB 9; Length 227;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      170 DSDALDSD 178
        |||||
DB      179 DSDALDSD 187

```

```

RESULT 3
US-10-424-599-234018
; Sequence 234018, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234018
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1.pep
US-10-424-599-234018

```

```

Query Match      1.4%; Score 9; DB 12; Length 324;
Best Local Similarity 100.0%; Pred. No. 5;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      224 ROMHCPACQ 232
        |||||
DB      119 ROMHCPACQ 127

```

RESULT 4

```

US-10-369-493-12141
; Sequence 12141, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12141
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-12141

```

```

Query Match      1.4%; Score 9; DB 15; Length 382;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      261 HRELAETVE 269
        |||||
DB      16 HRELAETVE 24

```

```

RESULT 5
US-10-425-114-58390
; Sequence 58390, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58390
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3079-030-A8_F11.pep
US-10-425-114-58390

```

```

Query Match      1.4%; Score 9; DB 12; Length 611;
Best Local Similarity 100.0%; Pred. No. 8.8;
Matches          9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      233 NGPGADMY 241
        |||||
DB      218 NGPGADMY 226

```

```

RESULT 6
US-10-282-122A-51413
; Sequence 51413, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos

```

```
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining prior application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 51413
LENGTH: 916
TYPE: PRT
ORGANISM: Bordetella pertussis
US-10-282-122A-51413

Query Match
Best Local Similarity 1.3%; Score 8; DB 12; Length 916;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 607 FDEALEQL 614
Db 654 FDEALEQL 661

RESULT 7
US-10-225-567A-1514
; Sequence 1514, Application US/10225567A
; Publication No. US20030113798A1
GENERAL INFORMATION:
APPLICANT: Lifespan Biosciences
APPLICANT: Brown, Joseph P.
APPLICANT: Burmer, Glenna C.
APPLICANT: Rouse, Christine L.
TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 1920-4-4
CURRENT APPLICATION NUMBER: US/10/225,567A
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/257,144
PRIOR FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 2292
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1514
LENGTH: 18
TYPE: PRT
ORGANISM: Homo sapiens
```

```
US-10-225-567A-1514
Query Match
Best Local Similarity 1.1%; Score 7; DB 14; Length 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 176 DSDDDLA 182
Db 3 DSDDDLA 9

RESULT 8
US-09-205-658-225
; Sequence 225, Application US/09205658
; Patent No. US20010029617A1
GENERAL INFORMATION:
APPLICANT: Ruvkun, Gary
APPLICANT: Osg, Scott
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
FILE REFERENCE: 00786/351004
CURRENT APPLICATION NUMBER: US/09/205,658
CURRENT FILING DATE: 1998-12-03
EARLIER APPLICATION NUMBER: 08/857,076
EARLIER FILING DATE: 1997-05-15
EARLIER APPLICATION NUMBER: 08/888,534
EARLIER FILING DATE: 1997-07-07
EARLIER APPLICATION NUMBER: US98/10080
NUMBER OF SEQ ID NOS: 328
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 225
LENGTH: 36
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-205-658-225

Query Match
Best Local Similarity 1.1%; Score 7; DB 9; Length 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 438 LRQISED 444
Db 2 LRQISED 8

RESULT 9
US-09-963-693-225
; Sequence 225, Application US/09963693
; Publication No. US20030181364A1
GENERAL INFORMATION:
APPLICANT: Ruvkun, Gary
APPLICANT: Osg, Scott
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
FILE REFERENCE: 00786/351004
CURRENT APPLICATION NUMBER: US/09/963,693
CURRENT FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US/09/205,658
PRIOR FILING DATE: 1998-12-03
PRIOR APPLICATION NUMBER: 08/857,076
PRIOR FILING DATE: 1997-05-15
PRIOR APPLICATION NUMBER: 08/888,534
PRIOR FILING DATE: 1997-07-07
PRIOR APPLICATION NUMBER: US98/10080
NUMBER OF SEQ ID NOS: 328
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 225
LENGTH: 36
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-963-693-225
```

Query Match 1.1%; Score 7; DB 10; Length 36;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 438 LRQISED 444  
 DB 2 LRQISED 8

## RESULT 10

US-10-424-599-250491  
 ; Sequence 250491, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 250491  
 ; LENGTH: 56  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_68221C.1.pcp  
 ; US-10-424-599-250491

Query Match 1.1%; Score 7; DB 12; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 277 ASVPCG 283  
 DB 31 ASVPCG 37

## RESULT 11

US-10-029-386-31679  
 ; Sequence 31679, Application US/10029386  
 ; Publication No. US20030194704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hantzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME ANALYSIS TWO  
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 ; FILE REFERENCE: AEMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 34288  
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 31679  
 ; LENGTH: 56  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AC008974.6  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.87  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.55  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8  
 ; OTHER INFORMATION: EXPRESSED IN  
 ; OTHER INFORMATION: SWISSPROT HIT: Q00535, EVALDB 2.80e+00  
 ; US-10-029-386-31679

Query Match 1.1%; Score 7; DB 14; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 370 ERLHREL 376  
 DB 47 ERLHREL 53

## RESULT 12

US-10-264-049-3845  
 ; Sequence 3845, Application US/10264049  
 ; Publication No. US20040005579A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Birse et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PA133PI  
 ; CURRENT APPLICATION NUMBER: US/10/264,049  
 ; CURRENT FILING DATE: 2002-10-04  
 ; PRIOR APPLICATION NUMBER: PCT/US01/18569  
 ; PRIOR FILING DATE: 2001-06-07  
 ; PRIOR APPLICATION NUMBER: US 60/209,467  
 ; PRIOR FILING DATE: 2000-06-07  
 ; NUMBER OF SEQ ID NOS: 4360  
 ; SOFTWARE: PatentIn Ver. 3.1  
 ; SEQ ID NO 3845  
 ; LENGTH: 57  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: MISC\_FEATURE  
 ; LOCATION: (1)  
 ; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
 ; US-10-264-049-3845

Query Match 1.1%; Score 7; DB 15; Length 57;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 543 QNINPS 549  
 DB 27 QNINPS 33

## RESULT 13

US-10-424-599-154923  
 ; Sequence 154923, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 154923  
 ; LENGTH: 60  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_110917C.1.pcp  
 ; US-10-424-599-154923

Query Match 1.1%; Score 7; DB 12; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 21 EYEQLVQ 27  
 DB 47 EYEQLVQ 53

## RESULT 14

```
US-10-424-599-228124
; Sequence 228124, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 228124
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(62)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_48023C.1.pep
; US-10-424-599-228124

Query Match          1.1%; Score 7; DB 12; Length 62;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      350 GHRGMSV 356
Db      30  GHRGMSV 36

RESULT 15
US-10-424-599-169910
; Sequence 169910, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 169910
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_124443C.1.pep
; US-10-424-599-169910

Query Match          1.1%; Score 7; DB 12; Length 68;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      561 SSFIEFO 567
Db      22  SSFIEFO 28
```

Search completed: April 6, 2004, 19:49:22  
Job time : 47 secs

**This Page Blank (uspio)**